

GENERAL PHYSICS

**Mechanics and
Molecular Physics**

L.D. Landau, A.I. Akhiezer,
E.M. Lifshitz

Pergamon Press

L. D. LANDAU, A. I. AKHIEZER
E. M. LIFSHITZ

GENERAL PHYSICS
Mechanics and Molecular Physics

Translated by

J. B. SYKES

A. D. PETFORD, C. L. PETFORD

PERGAMON PRESS

OXFORD · NEW YORK · TORONTO
SYDNEY · BRAUNSCHWEIG

Pure Mathematical Physics

Pergamon Press Ltd., Headington Hill Hall, Oxford

Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523

Pergamon of Canada Ltd., 207 Queen's Quay West, Toronto 1

Pergamon Press (Aust.) Pty. Ltd., 19a Boundary Street, Rushcutters Bay,
N.S.W. 2011, Australia

Vieweg & Sohn GmbH, Burgplatz 1, Braunschweig

Copyright © 1967 Pergamon Press Ltd.

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of Pergamon Press Ltd.

First English edition 1967

Reprinted 1970

Translated from *Kurs obshcheĭ fiziki: mekhanika i molekulyarnaya fizika*,
Izdatel'stvo "Nauka", Moscow, 1965

Library of Congress Catalog Card No. 67-30260

Filmset by Graphic Film Limited, Dublin
and printed in Great Britain by
J. W. Arrowsmith Limited, Bristol

0 08 009106 7

PREFACE

THE purpose of this book is to acquaint the reader with the principal phenomena and most important laws of physics. The authors have tried to make the book as compact as possible, including only what is essential and omitting what is of secondary significance. For this reason the discussion nowhere aims at anything approaching completeness.

The derivations of the formulae are given only in so far as they may help the reader in understanding the relations between phenomena. Formulae are therefore derived for simple cases wherever possible, on the principle that the systematic derivation of quantitative formulae and equations should rather appear in a textbook of theoretical physics.

The reader is assumed to be familiar with algebra and trigonometry and also to understand the fundamentals of the differential calculus and of vector algebra. He is further expected to have an initial knowledge of the main ideas of physics and chemistry. The authors hope that the book will be useful to physics students at universities and technical colleges, and also to physics teachers in schools.

This book was originally written in 1937, but has not been published until now, for a variety of reasons. It has now been augmented and rewritten, but the plan and essential content remain unchanged.

To our profound regret, L. D. Landau, our teacher and friend, has been prevented by injuries received in a road accident from personally contributing to the preparation of this edition. We have everywhere striven to follow the manner of exposition that is characteristic of him.

We have also attempted to retain as far as possible the original choice of material, being guided here both by the book in its original form and by the notes (published in 1948) taken from Landau's lectures on general physics in the Applied Physics Department of Moscow State University.

In the original plan, in order not to interrupt the continuity of the discussion, the methods of experimental study of thermal phenomena were to have been placed in a separate chapter at the end of the book. Unfortunately, we have not yet had an opportunity to carry out this intention, and we have decided, in order to avoid further delay, to publish the book without that chapter.

June 1965

A. I. AKHIEZER
E. M. LIFSHITZ

CHAPTER I

PARTICLE MECHANICS

§1. The principle of the relativity of motion

The fundamental concept of *mechanics* is that of *motion* of a body with respect to other bodies. In the absence of such other bodies it is clearly impossible to speak of motion, which is always relative. Absolute motion of a body irrespective of other bodies has no meaning.

The relativity of motion arises from the relativity of the concept of space itself. We cannot speak of position in absolute space independently of bodies therein, but only of position relative to certain bodies.

A group of bodies which are arbitrarily considered to be at rest, the motion of other bodies being taken as relative to that group, is called in physics a *frame of reference*. A frame of reference may be arbitrarily chosen in an infinite number of ways, and the motion of a given body in different frames will in general be different. If the frame of reference is the body itself, then the body will be at rest in that frame, while in other frames it will be in motion, and in different frames it will move differently, i.e. along different paths.

Different frames of reference are equally valid and equally admissible for investigating the motion of any given body. Physical phenomena, however, in general occur differently in different frames, and in this way different frames of reference may be distinguished.

It is reasonable to choose the frame of reference such that natural phenomena take their simplest form. Let us consider a body so far from other bodies that it does not interact with them. Such a body is said to be *moving freely*.

In reality, the condition of free motion can, of course, be fulfilled only to a certain approximation, but we can imagine in

principle that a body is free from interaction with other bodies to any desired degree of accuracy.

Free motion, like other forms of motion, appears differently in different frames of reference. If, however, the frame of reference is one in which any one freely moving body is fixed, then free motion of other bodies is especially simple: it is uniform motion in a straight line or, as it is sometimes called, motion with a velocity constant in magnitude and direction. This statement forms the content of the *law of inertia*, first stated by Galileo. A frame of reference in which a freely moving body is fixed is called an *inertial frame*. The law of inertia is also known as *Newton's first law*.

It might appear at first sight that the definition of an inertial frame as one with exceptional properties would permit a definition of absolute space and absolute rest relative to that frame. This is not so, in fact, since there exists an infinity of inertial frames: if a frame of reference moves with a velocity constant in magnitude and direction relative to an inertial frame, then it is itself an inertial frame.

It must be emphasised that the existence of inertial frames of reference is not purely a logical necessity. The assertion that there exist, in principle, frames of reference with respect to which the free motion of bodies takes place uniformly and in a straight line is one of the fundamental laws of Nature.

By considering free motion we evidently cannot distinguish between different inertial frames. It may be asked whether the examination of other physical phenomena might in some way distinguish one inertial frame from another and hence select one frame as having special properties. If this were possible, we could say that there is absolute space and absolute rest relative to this special frame of reference. There is, however, no such distinctive frame, since all physical phenomena occur in the same way in different inertial frames.

All the laws of Nature have the same form in every inertial frame, which is therefore physically indistinguishable from, and equivalent to, every other inertial frame. This result, one of the most important in physics, is called the *principle of relativity* of motion, and deprives of all significance the concepts of absolute space, absolute rest and absolute motion.

Since all physical laws are formulated in the same way in

every inertial frame, but in different ways in different non-inertial frames, it is reasonable to study any physical phenomenon in inertial frames, and we shall do so henceforward except where otherwise stated.

The frames of reference actually used in physical experiments are inertial only to a certain approximation. For example, the most usual frame of reference is that in which the Earth, on which we live, is fixed. This frame is not inertial, owing to the daily rotation of the Earth on its axis and its revolution round the Sun. These motions occur with different and varying velocities at different points on the Earth, and the frame in which the Earth is fixed is therefore not inertial. However, because of the relative slowness of variation of the direction of the velocities in the Earth's daily rotation on its axis and revolution round the Sun, we in fact commit a very small error, of no importance in many physical experiments, by assuming that the "terrestrial" frame of reference is an inertial frame. Although the difference between the motion in the terrestrial frame of reference and that in an inertial frame is very slight, it can nevertheless be observed, for example, by means of a Foucault pendulum, whose plane of oscillation slowly moves relative to the Earth's surface (§31).

§2. Velocity

It is reasonable to begin the study of the laws of motion by considering the motion of a body of small dimensions. The motion of such a body is especially simple because there is no need to take into account the rotation of the body or the relative movement of different parts of the body.

A body whose size may be neglected in considering its motion is called a *particle*, and is a fundamental object of study in mechanics. The possibility of treating the motion of a given body as that of a particle depends not only on its absolute size but also on the conditions of the physical problem concerned. For example, the Earth may be regarded as a particle in relation to its motion round the Sun, but not in relation to its daily rotation on its axis.

The position of a particle in space is entirely defined by specifying three coordinates, for instance the three Cartesian coordinates x , y , z . For this reason a particle is said to have three *degrees of freedom*. The quantities x , y , z form the *radius vector*

\mathbf{r} from the origin to the position of the particle.

The motion of a particle is described by its *velocity*. In uniform motion, the velocity is defined simply as the distance traversed by the particle in unit time. Generally, when the motion is not uniform and varies in direction, the particle velocity must be defined as a vector equal to the vector of an infinitesimal displacement ds of the particle divided by the corresponding infinitesimal time interval dt . Denoting the velocity vector by \mathbf{v} , we therefore have

$$\mathbf{v} = ds/dt.$$

The direction of the vector \mathbf{v} is the same as that of ds ; that is, the velocity at any instant is along the tangent to the path of the particle in the direction of motion.

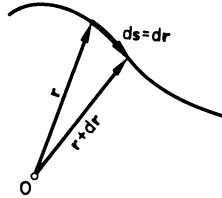


FIG. 1.

Figure 1 shows the path of a particle and the radius vectors \mathbf{r} and $\mathbf{r} + d\mathbf{r}$ at times t and $t + dt$. By the vector addition rule it is easily seen that the infinitesimal displacement ds of the particle is equal to the difference between the radius vectors at the initial and final instants: $ds = d\mathbf{r}$. The velocity \mathbf{v} may therefore be written

$$\mathbf{v} = d\mathbf{r}/dt,$$

and is thus the time derivative of the radius vector of the moving particle. Since the components of the radius vector \mathbf{r} are the coordinates x , y , z , the components of the velocity along these axes are the derivatives

$$v_x = dx/dt, \quad v_y = dy/dt, \quad v_z = dz/dt.$$

The velocity, like the position, is a fundamental quantity

describing the state of motion of a particle. The state of the particle is therefore defined by six quantities: three coordinates and three velocity components.

The relation between the velocities \mathbf{v} and \mathbf{v}' of the same particle in two different frames of reference K and K' may be determined as follows. If in a time dt the particle moves an amount ds relative to the frame K , and the frame K moves an amount dS relative to the frame K' , the vector addition rule shows that the displacement of the particle relative to the frame K' is $ds' = ds + dS$. Dividing both sides by the time interval dt and denoting the velocity of the frame K' relative to K by \mathbf{V} , we find

$$\mathbf{v}' = \mathbf{v} + \mathbf{V}.$$

This formula relating the velocities of a given particle in different frames of reference is called the *velocity addition rule*.

At first sight the velocity addition rule appears obvious, but in fact it depends on the tacitly made assumption that the passage of time is absolute. We have assumed that the time interval during which the particle moves by an amount ds in the frame K is equal to the time interval during which it moves by ds' in K' . In reality, this assumption proves to be not strictly correct, but the consequences of the non-absoluteness of time begin to appear only at very high velocities, comparable with that of light. In particular, the velocity addition rule is not obeyed at such velocities. In what follows we shall consider only velocities so small that the assumption of absolute time is quite justified.

The mechanics based on the assumption that time is absolute is called *Newtonian* or *classical mechanics*, and we shall here discuss only this mechanics. Its fundamental laws were stated by Newton in his *Principia* (1687).

§3. Momentum

In free motion of a particle, i.e. when it does not interact with other bodies, its velocity remains constant in any inertial frame of reference. If particles interact with one another, however, their velocities will vary with time; but the changes in the velocities of interacting particles are not completely independent of one another. In order to ascertain the nature of the relation between them, we define a *closed system*—a group of particles which

interact with one another but not with surrounding bodies. For a closed system there exist a number of quantities related to the velocities which do not vary with time. These quantities naturally play a particularly important part in mechanics.

One of these invariant or *conserved* quantities is called the *total momentum* of the system. It is the vector sum of the momenta of each of the particles forming a closed system. The momentum of a single particle is simply proportional to its velocity. The proportionality coefficient is a constant for any given particle and is called its *mass*. Denoting the particle momentum vector by \mathbf{p} and the mass by m , we can write

$$\mathbf{p} = m\mathbf{v},$$

where \mathbf{v} is the velocity of the particle. The sum of the vectors \mathbf{p} over all particles in the closed system is the total momentum of the system:

$$\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2 + \cdots = m_1\mathbf{v}_1 + m_2\mathbf{v}_2 + \cdots,$$

where the suffixes label the individual particles and the sum contains as many terms as there are particles in the system. This quantity is constant in time:

$$\mathbf{P} = \text{constant}.$$

Thus the total momentum of a closed system is conserved. This is the *law of conservation of momentum*. Since the momentum is a vector, the law of conservation of momentum separates into three laws expressing the constancy in time of the three components of the total momentum.

The law of conservation of momentum involves a new quantity, the mass of a particle. By means of this law, we can determine the ratios of particle masses. For let us imagine a collision between two particles of masses m_1 and m_2 , and let \mathbf{v}_1 and \mathbf{v}_2 denote their velocities before the collision, \mathbf{v}_1' and \mathbf{v}_2' their velocities after the collision. Then the law of conservation of momentum shows that

$$m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = m_1\mathbf{v}_1' + m_2\mathbf{v}_2'.$$

If Δv_1 and Δv_2 are the changes in the velocities of the two particles, this relation may be written as

$$m_1 \Delta v_1 + m_2 \Delta v_2 = 0,$$

whence

$$\Delta v_2 = -(m_1/m_2) \Delta v_1.$$

Thus the changes in velocity of two interacting particles are inversely proportional to their masses. Using this relation, we can find the ratio of the masses of the particles from the changes in their velocities. We must therefore arbitrarily take the mass of some particular body as unity and express the masses of all other bodies in terms of it. This unit of mass in physics is usually taken to be the gram (see §8).

§4. Motion under reactive forces

The law of conservation of momentum is one of the fundamental laws of Nature and plays a part in many phenomena. In particular, it accounts for motion under reactive forces.

We shall show how the velocity of a rocket may be found as a function of its varying mass. Let the velocity of the rocket relative to the Earth at some instant t be v and its mass M . At this instant, let the rocket begin to emit exhaust gases whose velocity relative to the rocket is u . In a time dt the mass of the rocket decreases to $M + dM$, where $-dM$ is the mass of the gas emitted, and the velocity increases to $v + dv$. Now let us equate the momenta of the system consisting of the rocket and the exhaust gases at times t and $t + dt$. The initial momentum is evidently Mv . The

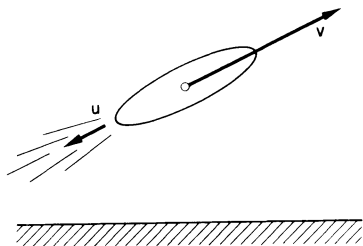


FIG. 2.

momentum of the rocket at time $t + dt$ is $(M + dM)(v + dv)$ (dM being negative) and the momentum of the exhaust gas is $-dM(v - u)$, since the velocity of the gas relative to the Earth is clearly $v - u$ (Fig. 2). The momenta at these two times must be equal, by the law of conservation of momentum:

$$Mv = (M + dM)(v + dv) - dM(v - u),$$

whence, neglecting the second-order small quantity $dM dv$, we obtain

$$M dv + u dM = 0$$

or

$$dM/M = -dv/u.$$

We shall suppose that the gas outflow velocity does not vary with time. Then the last equation may be written

$$d \log_e M = -d(v/u),$$

and therefore

$$\log_e M + v/u = \text{constant}.$$

The value of the constant is given by the condition that the mass of the rocket is M_0 at the beginning of its motion (i.e. when $v = 0$), so that

$$\log_e M + v/u = \log_e M_0,$$

whence we have finally

$$v = u \log_e(M_0/M).$$

This formula gives the velocity of the rocket as a function of its varying mass.

§5. Centre of mass

The law of conservation of momentum is related to an important property of mass, the *law of conservation of mass*. In order to understand the meaning of this law, let us consider the point

called the *centre of mass* of a closed system of particles. The coordinates of the centre of mass are the mean values of the coordinates of the particles, the coordinate of each particle being counted as many times as its mass exceeds the unit mass. That is, if x_1, x_2, \dots , denote the x coordinates of particles having masses m_1, m_2, \dots , then the x coordinate of the centre of mass is determined by the formula

$$X = \frac{m_1x_1 + m_2x_2 + \dots}{m_1 + m_2 + \dots}.$$

Similar formulae may be written for the y and z coordinates. These formulae can be put in a single vector form as an expression for the radius vector \mathbf{R} of the centre of mass:

$$\mathbf{R} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + \dots}{m_1 + m_2 + \dots},$$

where $\mathbf{r}_1, \mathbf{r}_2, \dots$ are the radius vectors of the individual particles.

The centre of mass has the noteworthy property of moving with constant velocity, whereas the individual particles forming the closed system may move with velocities which vary with time. For the velocity of the centre of mass is

$$\mathbf{V} = \frac{d\mathbf{R}}{dt} = \frac{m_1d\mathbf{r}_1/dt + m_2d\mathbf{r}_2/dt + \dots}{m_1 + m_2 + \dots}.$$

But $d\mathbf{r}_1/dt$ is the velocity of the first particle, $d\mathbf{r}_2/dt$ that of the second particle, and so on. Denoting these by $\mathbf{v}_1, \mathbf{v}_2, \dots$, we have

$$\mathbf{V} = \frac{m_1\mathbf{v}_1 + m_2\mathbf{v}_2 + \dots}{m_1 + m_2 + \dots}.$$

The numerator is the total momentum of the system, which we have denoted by \mathbf{P} , and we therefore have finally

$$\mathbf{V} = \mathbf{P}/M,$$

where M is the total mass of the particles: $M = m_1 + m_2 + \dots$. Since the total momentum of the system is conserved, the velocity of the centre of mass is constant in time.

Writing this formula as

$$\mathbf{P} = M\mathbf{V},$$

we see that the total momentum of the system, the velocity of its centre of mass and the total mass of the particles in the system are related in the same way as the momentum, velocity and mass of a single particle. We can regard the total momentum of the system as the momentum of a single particle at the centre of mass of the system, with a mass equal to the total mass of the particles in the system. The velocity of the centre of mass may be regarded as the velocity of the system of particles as a whole, and the sum of the individual masses appears as the mass of the whole system.

Thus we see that the mass of a composite body is equal to the sum of the masses of its parts. This is a very familiar assertion and might appear to be self-evident; but it is in fact by no means trivial and represents a physical law which follows from the law of conservation of momentum.

Since the velocity of the centre of mass of a closed system of particles is constant in time, a frame of reference in which the centre of mass is fixed is an inertial frame, called the *centre-of-mass frame*. The total momentum of a closed system of particles is obviously zero in this frame. The description of phenomena in this frame of reference eliminates complications arising from the motion of the system as a whole, and demonstrates more clearly the properties of the internal processes occurring within the system. For this reason the centre-of-mass frame is frequently used in physics.

§6. Acceleration

For a particle moving in a general manner the velocity varies continually in both magnitude and direction. Let the velocity change by dv in a time dt . If the change per unit time is taken we have the *acceleration* vector of the particle, denoted here by \mathbf{w} :

$$\mathbf{w} = dv/dt.$$

Thus the acceleration determines the change in the velocity of the particle and is equal to the time derivative of the velocity.

If the direction of the velocity is constant, i.e. the particle moves in a straight line, then the acceleration is along that line and is clearly

$$w = dv/dt.$$

It is also easy to determine the acceleration when the velocity remains constant in magnitude but varies in direction. This case occurs when a particle moves uniformly in a circle.

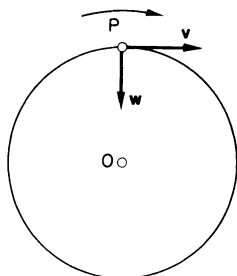


FIG. 3.

Let the velocity of the particle at some instant be v (Fig. 3). We mark v from a point C on an auxiliary diagram (Fig. 4). When the particle moves uniformly in a circle, the end of the vector v (the point A) also moves uniformly in a circle of radius v equal to the magnitude of the velocity. It is clear that the velocity of A will be equal to the acceleration of the original particle P , since the motion of A in a time dt is dv and its velocity is therefore dv/dt . This velocity is tangential to the circle round C and is perpendicular to v ; in the diagram it is shown by w . If we draw the vector w at the point P it will obviously be directed towards the centre O of the circle.

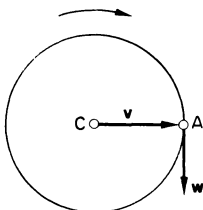


FIG. 4.

Thus the acceleration of a particle moving uniformly in a circle is towards the centre of the circle, i.e. at right angles to the velocity.

Let us now determine the magnitude of the acceleration w . To do so, we must find the velocity of the point A moving in a circle of radius v . When P moves once round the circle about O , in a time T , say, the point A traverses the circle about C , a distance $2\pi v$. The velocity of A is therefore

$$w = 2\pi v/T.$$

Substituting the period $T = 2\pi r/v$, where r is the radius of the path of the particle P , we obtain finally

$$w = v^2/r.$$

Thus, if the velocity varies only in magnitude, the acceleration is parallel to the velocity; if the velocity varies only in direction, the acceleration and velocity vectors are mutually perpendicular.

In general, when the velocity varies in both magnitude and direction, the acceleration has two components, one parallel to the velocity and one perpendicular to it. The first component, called the *tangential* component, is equal to the time derivative of the magnitude of the velocity:

$$w_t = dv/dt.$$

The second component, w_n , is called the *normal* component of the acceleration. It is proportional to the square of the velocity of the particle and inversely proportional to the radius of curvature of the path at the point considered.

§7. Force

If a particle is in free motion, i.e. does not interact with surrounding bodies, its momentum is conserved. If, on the other hand, the particle interacts with surrounding bodies, then its momentum varies with time. We can therefore regard the change in momentum of a particle as a measure of the action of surrounding bodies on it. The greater the change (per unit time), the stronger the action. It is therefore reasonable to take the time

derivative of the momentum vector of the particle in order to define this action. The time derivative is called the *force* on the particle.

This definition describes one aspect of the interaction: it concerns the extent of the “reaction” of the particle to the action of surrounding bodies on it. Conversely, by studying the interaction of the particle with surrounding bodies, we can relate the strength of this interaction to quantities describing the state of the particle and that of the surrounding bodies.

The forces of interaction between particles depend (in classical mechanics) only on their position. In other words, the forces acting between particles depend only on the distances between them and not on their velocities.

The manner in which the forces depend on the distances between the particles can in many cases be established by an examination of the physical phenomena underlying the interaction between particles.

Let \mathbf{F} denote the force acting on a given particle, expressed as a function of its coordinates and of quantities representing the properties and positions of the surrounding bodies. We can then write down an equation between two expressions for the force: \mathbf{F} , and the change in the momentum \mathbf{p} of the particle per unit time,

$$d\mathbf{p}/dt = \mathbf{F}.$$

This is called the *equation of motion* of the particle.

Since $\mathbf{p} = m\mathbf{v}$, the equation of motion of a particle may also be written

$$m d\mathbf{v}/dt = \mathbf{F}.$$

Thus the force acting on a particle is equal to the product of its acceleration and its mass. This is *Newton's second law*.

It should be emphasised, however, that this law acquires a specific significance only when \mathbf{F} is known as a function of the particle coordinates. In that case, i.e. if the form of the function \mathbf{F} is known, the equation of motion enables us, in principle, to determine the velocity and coordinates of the particle as functions of time; that is, to find its path. In addition to the form of the function \mathbf{F} (i.e. the law of interaction between the particle and

surrounding bodies), the *initial conditions* must be given, that is, the position and velocity of the particle at some instant taken as the initial instant. Since the equation of motion determines the increment of velocity of the particle in any time interval dt ($d\mathbf{v} = \mathbf{F} dt/m$), and the velocity gives the change in spatial position of the particle ($d\mathbf{r} = \mathbf{v} dt$), it is clear that specifying the initial position and initial velocity of the particle is in fact sufficient to determine its further motion completely. This is the significance of the statement made in §2 that the mechanical state of a particle is defined by its coordinates and velocity.

The equation of motion is a vector equation, and may therefore be written as three equations relating the components of acceleration and force:

$$m dv_x/dt = F_x, \quad m dv_y/dt = F_y, \quad m dv_z/dt = F_z.$$

Let us now consider a closed system of particles. As we know, the sum of the momenta of such particles is conserved:

$$\mathbf{p}_1 + \mathbf{p}_2 + \cdots = \text{constant},$$

where \mathbf{p}_i is the momentum of the i th particle. Differentiation of this equation with respect to time gives

$$\frac{d\mathbf{p}_1}{dt} + \frac{d\mathbf{p}_2}{dt} + \cdots = 0.$$

Since

$$d\mathbf{p}_i/dt = \mathbf{F}_i,$$

where \mathbf{F}_i is the force on the i th particle, we have

$$\mathbf{F}_1 + \mathbf{F}_2 + \cdots = 0.$$

Thus the sum of all the forces in a closed system is zero.

In particular, if the closed system contains only two bodies, the force exerted by one body on the other must be equal in magnitude and opposite in direction to the force which the latter body exerts on the former. This is called the *law of action and*



FIG. 5.

reaction or *Newton's third law*. Since, in this case, there is only one distinctive direction, namely that of the line joining the bodies (or particles), the forces F_1 and F_2 must act along this line (see Fig. 5, where M_1 and M_2 denote the two particles).

§8. Dimensions of physical quantities

All physical quantities are measured in certain units. To measure a quantity is to determine its ratio to another quantity of the same kind which is arbitrarily taken as the unit.

In principle, any unit may be chosen for each physical quantity, but by using the relations between different quantities it is possible to define a limited number of arbitrary units for certain quantities taken as fundamental, and to construct for the other quantities units which are related to the fundamental units. These are called *derived* units.

Length, time and mass are taken as the fundamental quantities in physics.

The unit of length in physics is the centimetre (cm), equal to one-hundredth of the metre, which is now defined as equal to 1 650 763·73 wavelengths of the light corresponding to a particular (orange) line in the spectrum of the gas krypton.

The metre was originally defined as one ten-millionth of a quadrant of the meridian through Paris, and the standard metre was constructed from measurements made in 1792. Since it was extremely difficult to reproduce the standard metre on the basis of its "natural" definition, the metre was later defined by agreement as the length of a particular standard—a platinum-iridium bar preserved at the International Bureau of Weights and Measures in Paris. This definition of the metre as a "distance between lines" has also now been abandoned, and the "light" metre described above is used. In consequence, the unit of length is again a natural and indestructible measure of length, and, moreover, allows a hundredfold increase in the accuracy of reproduction of the standard metre.

The following units are used in measuring short distances: the *micron* ($1 \mu = 10^{-4}$ cm), the *millimicron* ($1 m\mu = 10^{-7}$ cm), the *ångström* ($1 \text{ \AA} = 10^{-8}$ cm) and the *fermi* (10^{-13} cm).

In astronomy, distances are measured in terms of the *light-year*, the distance traversed by light in one year, equal to 9.46×10^{17} cm. A distance of 3.25 light-years or 3.08×10^{18} cm is called a *parsec*; it is the distance at which the diameter of the Earth's orbit subtends an angle of one second of arc.

Time in physics is measured in seconds. The second (sec) is now defined as a certain fraction of a particular tropical year (1900). The tropical year is the time between successive passages of the Sun through the vernal equinox. The year (1900) is specified because the length of the tropical year is not constant but decreases by about 0.5 sec per century.

The second was originally defined as a fraction ($1/86\,400$) of the solar day, but the Earth's daily rotation is not uniform and the length of the day is not constant. The relative fluctuations of the length of the day are about 10^{-7} , which is too great for the day to be used as a basis for the definition of the unit of time, in terms of present-day technology. The relative fluctuations in the length of the tropical year are smaller, but the definition of the second on the basis of the Earth's revolution round the Sun cannot be regarded as entirely satisfactory, since it does not allow a "standard" unit of time to be reproduced with sufficient accuracy. This difficulty disappears only if the definition of the second is based not on the motion of the Earth but on the periodic motions occurring in atoms. The second then becomes a natural physical unit of time just as the "light" centimetre is a natural unit of length.

Mass in physics is measured in grams, as already stated. One gram (g) is one-thousandth of the mass of a standard kilogram preserved at the International Bureau of Weights and Measures in Paris.

The mass of one kilogram was originally defined as the mass of one cubic decimetre of water at 4°C , i.e. the temperature at which water has its maximum density. It was, however, impossible to maintain this definition, as with the original definition of the metre, owing to the increasing accuracy of measurements; if the original definitions were retained it would be necessary to keep changing the fundamental standards. Modern results show that 1 cm^3 of distilled water at 4°C weighs not 1 g but 0.999 972 g.

The definition of the kilogram as the mass of a standard, however, suffers from the same defects as the definition of the metre

as a “distance between lines”. The best procedure would be to define the gram not in terms of the mass of a standard kilogram but in terms of the mass of an atomic nucleus, such as the proton.

Let us now see how derived units are constructed, taking a few examples.

As the unit of velocity we could take any arbitrary velocity (for instance, the mean velocity of the Earth round the Sun, or the velocity of light), and refer all other velocities to this as the unit; but we can also use the definition of velocity as the ratio of distance to time and take as the unit of velocity the velocity at which a distance of one centimetre is traversed in one second. This velocity is denoted by 1 cm/sec. The symbol cm/sec is called the *dimensions* of velocity in terms of the fundamental units, the centimetre for length and the second for time. The dimensions of velocity are written

$$[v] = \text{cm/sec.}$$

The situation is similar for acceleration. The unit of acceleration could be taken to be any acceleration (for instance, that of a freely falling body), but we can use the definition of acceleration as the change of velocity per unit time, and take as the unit of acceleration the acceleration such that the velocity changes by 1 cm/sec in one second. The notation for this unit is 1 cm/sec², and the symbol cm/sec² denotes the dimensions of acceleration, written as

$$[w] = \text{cm/sec}^2.$$

Let us now determine the dimensions of force and establish the unit of force. To do so, we use the definition of force as the product of mass and acceleration. Using square brackets to denote the dimensions of any physical quantity, we obtain for the dimensions of force the expression

$$[F] = [m][w] = \text{g.cm/sec}^2.$$

As the unit of force we can take 1 g.cm/sec², which is called one *dyne*. This is the force which gives a mass of 1 g an acceleration of 1 cm/sec².

Thus, by using the relations between various quantities, we can choose units for all physical quantities by starting from a small number of fundamental quantities whose units are chosen arbitrarily. The system of physical units with the centimetre, gram and second as the fundamental units of mass, length and time is called the *physical* or *CGS system of units*.

It should not be thought, however, that the use of just three arbitrary fundamental units in this system has any deep physical significance. It arises only from the practical convenience of the system constructed from these units. In principle, a system of units could be constructed with any number of arbitrarily chosen units (see §22).

Operations with dimensions are carried out as if the latter were ordinary algebraic quantities, i.e. they are subject to the same operations as numbers. The dimensions of both sides of any equation containing different physical quantities must obviously be the same. This fact should be remembered in checking formulae.

It is often known from physical considerations that a particular physical quantity can depend only on certain other quantities. In many cases dimensional arguments alone suffice for the nature of the dependence to be determined. We shall later see examples of this.

Besides the CGS system of units, other systems are frequently used, in which the fundamental units of mass and length are greater than the gram and the centimetre. The international system of units (SI) is based on the metre, kilogram and second as units of length, mass and time. The unit of force in this system is called the *newton* (N):

$$1 \text{ N} = 1 \text{ kg.m/sec}^2 = 10^5 \text{ dyn.}$$

In engineering calculations, force is usually measured in units of kilogram-force (kgf). This is the force with which a mass of 1 kg is attracted to the Earth at sea level in latitude 45° . Its value is

$$1 \text{ kgf} = 9.8 \times 10^5 \text{ dyn} = 9.8 \text{ N}$$

(more precisely, 980 665 dyn).

§9. Motion in a uniform field

If a particle is subject to a definite force at every point in space, these forces as a whole are called a *force field*. In general, the field forces may vary from one point to another in space and may also depend on time.

Let us consider the simple case of the motion of a particle in a *uniform and constant field*, where the field forces have the same magnitude and direction everywhere and are independent of time, for example the Earth's gravitational field in regions small compared with its radius.

From the equation of motion of a particle,

$$m \, d\mathbf{v}/dt = \mathbf{F},$$

we have when \mathbf{F} is constant

$$\mathbf{v} = (\mathbf{F}/m)t + \mathbf{v}_0,$$

where \mathbf{v}_0 is the initial velocity of the particle. Thus in a uniform and constant field the velocity is a linear function of time.

The expression obtained for \mathbf{v} shows that the particle moves in the plane defined by the force vector \mathbf{F} and the initial velocity vector \mathbf{v}_0 . Let us take this as the xy plane, and the y axis in the direction of the force \mathbf{F} . The equation for the velocity \mathbf{v} of the particle gives two equations for the velocity components v_x and v_y :

$$v_y = (F/m)t + v_{y0}, \quad v_x = v_{x0},$$

where v_{x0} and v_{y0} are the initial values of the velocity components.

Since the velocity components are the time derivatives of the corresponding coordinates of the particle, we can write the last two equations as

$$dy/dt = (F/m)t + v_{y0}, \quad dx/dt = v_{x0}.$$

Hence

$$y = (F/2m)t^2 + v_{y0}t + y_0,$$

$$x = v_{x0}t + x_0,$$

where x_0 and y_0 are the initial values of the coordinates of the particle. These expressions determine the path of the particle. They can be simplified if time is measured from the instant at which the velocity component v_y is zero; then $v_{y0} = 0$. Taking the

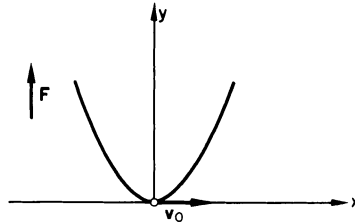


FIG. 6.

origin at the point where the particle is at that instant, we have $x_0 = y_0 = 0$. Finally, denoting the quantity v_{x0} , which is now the initial magnitude of the velocity, by v_0 simply, we have

$$y = (F/2m)t^2, \quad x = v_0 t.$$

Elimination of t gives

$$y = (F/2mv_0^2)x^2,$$

the equation of a parabola (Fig. 6). Thus a particle in a uniform field describes a parabola.

§10. Work and potential energy

Let us consider the motion of a particle in a force field \mathbf{F} . If the particle moves an infinitesimal distance ds under the action of the force \mathbf{F} , the quantity

$$dA = F ds \cos \theta,$$

where θ is the angle between the vectors \mathbf{F} and ds , is called the *work* done by the force \mathbf{F} over the distance ds . The product of the magnitudes of two vectors \mathbf{a} and \mathbf{b} and the cosine of the angle between them is called the *scalar product* of these vectors and denoted by $\mathbf{a} \cdot \mathbf{b}$. The work may therefore be defined as the scalar product of the force vector and the particle displacement vector:

$$dA = \mathbf{F} \cdot ds.$$

This expression may also be written

$$dA = F_s ds,$$

where F_s is the component of the force \mathbf{F} in the direction of motion of the particle.

In order to determine the work done by field forces over a finite path of the particle, it is necessary to divide this path into infinitesimal intervals ds , find the work for each such interval, and add the results. The sum gives the work done by the field forces over the whole path.

From the definition of work it follows that a force perpendicular to the path does no work. In particular, in uniform motion of a particle in a circle the work done by forces is zero.

A constant force field, i.e. one independent of time, has the following remarkable property: if a particle moves along a closed path in such a field, so as to return to its original position, then the work done by the field forces is zero.

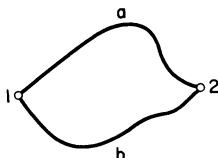


FIG. 7.

From this property there follows another result: the work done by the field forces in moving a particle from one position to another is independent of the path taken, and is determined only by the initial and final points. For let us consider two points 1 and 2 joined by two curves a and b (Fig. 7), and suppose that the particle moves from point 1 to point 2 along curve a and then from point 2 back to point 1 along curve b . The total work done by the field forces during this process is zero. Denoting the work by A , we can write

$$A_{1a2} + A_{2b1} = 0.$$

When the direction of motion is reversed, the sign of the work is obviously changed, and thus we have

$$A_{1a2} = -A_{2b1} = A_{1b2},$$

i.e. the work is independent of the form of the curve joining the initial and final points 1 and 2.

Since the work done by the field forces is independent of the path taken and is determined only by the terminal points of the path, it is clearly a quantity of deep physical significance. It can be used to define an important property of the force field. To do so, we take any point O in space as the origin, and consider the work done by the field forces when the particle moves from O to any point P , denoting this work by $-U$. The quantity U , i.e. minus the work done in moving the particle from O to P , is called the *potential energy* of the particle at the point P . It is a function of the coordinates x, y, z of the point P :

$$U = U(x, y, z).$$

The work A_{12} done by the field forces when the particle moves between any points 1 and 2 is

$$A_{12} = U_1 - U_2,$$

where U_1 and U_2 are the values of the potential energy at the two points. The work done is equal to the difference of the potential energies at the initial and final points of the path.

Let us consider two points P and P' an infinitesimal distance apart. The work done by the field forces when the particle moves from P to P' is $-dU$. This work is also equal to $\mathbf{F} \cdot d\mathbf{s}$, where $d\mathbf{s}$ is the vector from P to P' ; it has been shown in §2 that the vector $d\mathbf{s}$ is equal to the difference $d\mathbf{r}$ of the radius vectors of P' and P . Thus we obtain the equation

$$\mathbf{F} \cdot d\mathbf{r} = -dU.$$

This relation between the force and the potential energy is one of the fundamental relations of mechanics.

Writing $\mathbf{F} \cdot d\mathbf{r} = \mathbf{F} \cdot d\mathbf{s} = F_s ds$, we can put this relation in the form

$$F_s = -dU/ds.$$

This means that the component of the force in any direction is obtained by dividing the infinitesimal change dU in the potential

energy over an infinitesimal interval in that direction by the length ds of the interval. The quantity dU/ds is called the derivative of U in the direction s .

To explain these relations, let us determine the potential energy in a constant uniform field. We take the direction of the field force F as the z axis. Then $\mathbf{F} \cdot d\mathbf{r} = Fdz$; equating this to the change in the potential energy, we have $-dU = Fdz$, whence

$$U = -Fz + \text{constant.}$$

We see that the potential energy is defined only to within an arbitrary constant. This is a general result related to the arbitrariness of the choice of the original point O in the field from which the work done on the particle is measured. It is usual to choose the arbitrary constant in the expression for U so that the potential energy of the particle is zero when it is at an infinite distance from other bodies.

From the relations between the force components and the potential energy we can deduce the direction of the force. If the potential energy increases in a given direction ($dU/dt > 0$), the component of the force in that direction is negative, i.e. the force is in the direction of decreasing potential energy. Force always acts in the direction in which potential energy decreases.

Since the derivative vanishes at points where the function has a maximum or minimum, the force is zero at points of maximum or minimum potential energy.

§11. The law of conservation of energy

The fact that the work done by the forces of a constant field when a particle moves from one point to another is independent of the shape of the path along which the particle moves leads to an extremely important relationship, the law of conservation of energy.

In order to derive this, we recall that the force \mathbf{F} acting on the particle is

$$\mathbf{F} = m d\mathbf{v}/dt.$$

Since the component of acceleration in the direction of the motion is dv/dt , the force component in this direction is

$$F_s = m dv/dt.$$

Let us now determine the work done by this force over an infinitesimal distance $ds = v dt$:

$$dA = F_s ds = mv dv,$$

or

$$dA = d(\frac{1}{2}mv^2).$$

Thus the work done by the force is equal to the increase in $\frac{1}{2}mv^2$. This quantity is called the *kinetic energy* of the particle.

The work is also equal to the decrease in potential energy: $dA = -dU$. We can therefore write

$$-dU = d(\frac{1}{2}mv^2),$$

i.e.

$$d(U + \frac{1}{2}mv^2) = 0.$$

Denoting the sum by E , we hence obtain

$$E = \frac{1}{2}mv^2 + U = \text{constant}.$$

Thus the sum of the kinetic energy of the particle, which depends only on its velocity, and the potential energy, which depends only on its coordinates, is constant during the motion of the particle. This sum is called the total energy or simply the *energy* of the particle, and the relationship derived above is called the *law of conservation of energy*.

The force field in which the particle moves is generated by various other bodies. If the field is constant, these bodies must be at rest. Thus we have derived the law of conservation of energy in the simple case where one particle moves and all the other bodies with which it interacts are at rest. But the law of conservation of energy can also be stated in the general case where more than one particle is moving. If these particles form a closed system, a law of conservation of energy is again valid which states that the sum of the kinetic energies of all the particles

separately and their mutual potential energy does not vary with time, i.e.

$$E = \frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2 + \dots + U(\mathbf{r}_1, \mathbf{r}_2, \dots),$$

where m_i is the mass of the i th particle, v_i its velocity and U the potential energy of interaction of the particles, which depends on their radius vectors \mathbf{r}_i .

The function U is related to the forces acting on each particle in the same way as for a single particle in an external field. In determining the force \mathbf{F}_i acting on the i th particle we must consider the change in the potential energy U in an infinitesimal displacement $d\mathbf{r}_i$ of this particle, the positions of all the other particles remaining unchanged. The work $\mathbf{F}_i \cdot d\mathbf{r}_i$ done on the particle in such a displacement is equal to the corresponding decrease in the potential energy.

The law of conservation of energy is valid for any closed system and, like the law of conservation of momentum, is one of the most important laws of mechanics.

The kinetic energy is an essentially positive quantity. The potential energy of interaction of particles may be either positive or negative. If the potential energy of two particles is so defined that it is zero when the particles are at a great distance apart, its sign depends on whether the interaction between the particles is attractive or repulsive. Since the forces acting on particles are always in the direction of decreasing potential energy, the approach of attracting particles leads to a decrease in potential energy, which is therefore negative. The potential energy of repelling particles, on the other hand, is positive.

Energy, and also work, have the dimensions

$$[E] = [m][v^2] = \text{g.cm}^2/\text{sec}^2.$$

The unit of energy in the CGS system is therefore $1 \text{ g.cm}^2/\text{sec}^2$, which is called the *erg*. It is the work done by a force of 1 dyn acting through 1 cm.

In the SI system a larger unit of energy, the *joule* (J), is used, equal to the work done by a force of 1 N acting through 1 m:

$$1 \text{ J} = 1 \text{ N.m} = 10^7 \text{ erg}.$$

If the unit of force is the kilogram-force, the corresponding unit of energy is the kilogram-metre (kgf.m), equal to the work done by a force of 1 kgf acting through 1 m. It is related to the joule by $1 \text{ kgf.m} = 9.8 \text{ J}$.

Energy sources are described by the work done per unit time. This is called the *power*. The unit of power is the *watt* (W):

$$1 \text{ W} = 1 \text{ J/sec.}$$

The work done in one hour by an energy source of power 1 W is called a *watt-hour* (Wh). It is easy to see that

$$1 \text{ Wh} = 3.6 \times 10^3 \text{ J.}$$

§12. Internal energy

As has been explained in §5, for the motion of a composite system we can define the velocity of the system as a whole, namely the velocity of the centre of mass of the system. This means that the motion of the system may be regarded as consisting of two parts: the motion of the system as a whole and the “internal” motion of the particles forming the system relative to the centre of mass. Accordingly the energy E of the system may be written as the sum of the kinetic energy of the system as a whole, which is $\frac{1}{2}MV^2$ (where M is the mass of the system and V the velocity of its centre of mass), and the *internal energy* E_{int} of the system, which comprises the kinetic energy of the internal motion of the particles and the potential energy of their interaction:

$$E = \frac{1}{2}MV^2 + E_{\text{int}}.$$

Although this formula is fairly obvious, we shall also give a direct derivation of it. The velocity of the i th particle, say, relative to a fixed frame of reference may be written as $\mathbf{v}_i + \mathbf{V}$, where \mathbf{V} is the velocity of the centre of mass of the system and \mathbf{v}_i is the velocity of the particle relative to the centre of mass. The kinetic energy of the particle is

$$\frac{1}{2}m_i(\mathbf{v}_i + \mathbf{V})^2 = \frac{1}{2}m_iV^2 + \frac{1}{2}m_iv_i^2 + m_i\mathbf{V} \cdot \mathbf{v}_i.$$

On summation over all particles the first term from each such expression gives $\frac{1}{2}MV^2$, where $M = m_1 + m_2 + \dots$. The sum of

the second terms gives the total kinetic energy of the internal motion in the system. The sum of the third terms is zero, since

$$m_1 \mathbf{V} \cdot \mathbf{v}_1 + m_2 \mathbf{V} \cdot \mathbf{v}_2 + \cdots = \mathbf{V} \cdot (m_1 \mathbf{v}_1 + m_2 \mathbf{v}_2 + \cdots);$$

the expression in parentheses is the total momentum of the particles relative to the centre of mass of the system, which by definition is zero. Finally, adding the kinetic energy to the potential energy of interaction of the particles, we obtain the required formula.

Using the law of conservation of energy, we can discuss the stability of a composite body. The problem here is to ascertain the conditions in which the composite body may spontaneously disintegrate into its component parts. Let us consider, for example, the break-up of a composite body into two parts; let the masses of the parts be m_1 and m_2 , and let the velocities of the parts in the centre-of-mass frame of the original composite body be \mathbf{v}_1 and \mathbf{v}_2 . Then the law of conservation of energy in this frame is

$$E_{\text{int}} = \frac{1}{2}m_1v_1^2 + E_{1\text{int}} + \frac{1}{2}m_2v_2^2 + E_{2\text{int}},$$

where E_{int} is the internal energy of the original body and $E_{1\text{int}}$, $E_{2\text{int}}$ the internal energies of the two parts. Since the kinetic energy is always positive, it follows from the above relation that

$$E_{\text{int}} > E_{1\text{int}} + E_{2\text{int}}.$$

This is the condition for the body to be able to disintegrate into two parts. If, on the other hand, the internal energy of the body is less than the sum of the internal energies of its component parts, the body will be stable with respect to the disintegration.

§13. Boundaries of the motion

If the motion of a particle is constrained so that it can move only along a certain curve, the motion is said to have one degree of freedom or to be *one-dimensional*. One coordinate is then sufficient to specify the position of the particle; it may be taken, for example, as the distance along the curve from a point taken as origin. Let this coordinate be denoted by x . The potential energy

of a particle in one-dimensional motion is a function only of this one coordinate: $U = U(x)$.

According to the law of conservation of energy we have

$$E = \frac{1}{2}mv^2 + U(x) = \text{constant},$$

and since the kinetic energy cannot take negative values the inequality

$$U \leq E$$

must hold. This implies that the particle during its motion can occupy only points where the potential energy does not exceed the total energy. If these energies are equal, we have the equation

$$U(x) = E,$$

which determines the limiting positions of the particle.

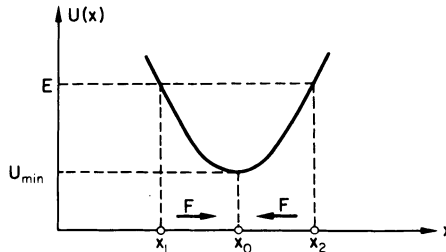


FIG. 8.

Some typical examples are the following. Let us first take a potential energy which, as a function of the coordinate x , has the form shown in Fig. 8. In order to find the boundaries of the motion of a particle in such a force field, as functions of the total energy E of the particle, we draw a straight line $U = E$ parallel to the x axis. This line intersects the curve of potential energy $U = U(x)$ at two points, whose abscissae are denoted by x_1 and x_2 . If the motion is to be possible it is necessary that the potential energy should not exceed the total energy. This means that the motion of a particle with energy E can occur only between the points x_1 and x_2 , and a particle of energy E cannot enter the regions right of x_2 and left of x_1 .

A motion in which the particle remains in a finite region of space is called a *finite motion*; one in which the particle can go to any distance is called an *infinite motion*.

The region of finite motions depends, of course, on the energy; in the example considered here, it decreases with decreasing energy and shrinks to a single point x_0 when $E = U_{\min}$.

At the points x_1 and x_2 the potential energy is equal to the total energy, and therefore at these points the kinetic energy and hence the particle velocity are zero. At the point x_0 the potential energy is a minimum, and the kinetic energy and velocity have their maximum values. Since the force F is related to the potential energy $F = -dU/dx$, it is negative between x_0 and x_2 , and positive between x_0 and x_1 . This means that between x_0 and x_2 the force is in the direction of decreasing x , i.e. to the left, and between x_0 and x_1 it is to the right. Consequently, if the particle begins to move from the point x_1 , where its velocity is zero, the force to the right will gradually accelerate it to a maximum velocity at the point x_0 . As the particle continues to move from x_0 to x_2 under the force which is now to the left, it will slow down until it comes to rest at x_2 . It will then begin to move back from x_2 to x_0 . This type of motion will continue indefinitely. Thus the particle executes a periodic motion with a period equal to twice the time for the particle to go from x_1 to x_2 .

At the point x_0 the potential energy is a minimum and the derivative of U with respect to x is zero; at this point the force is therefore zero, and the point x_0 is consequently a position of equilibrium of the particle. This position is evidently one of *stable equilibrium*, since in this case a departure of the particle from the equilibrium position causes a force which tends to return the particle to the equilibrium position. This property exists only for minima and not for maxima of the potential energy, although at the latter the force is likewise zero. If a particle is moved in either direction from a point of maximum potential energy, the resulting force in either case acts away from this point, and points where the potential energy reaches a maximum are therefore positions of *unstable equilibrium*.

Let us now consider the motion of a particle in a more complex field whose potential-energy curve has the form shown in Fig. 9. This curve has both a minimum and a maximum. If the particle has energy E , it can move in such a field in two regions: region I

between the points x_1 and x_2 , and region III to the right of the point x_3 (at these points the potential energy is equal to the total energy). The motion in the former region is of the same type as in the previous example, and is oscillatory. The motion in region III, however, is infinite, since the particle may move to any distance to the right of the point x_3 . If the particle begins its motion at the point x_3 , where its velocity is zero, it will continually be accelerated by the force to the right; at infinity, the potential energy is zero and the particle velocity reaches the value $v_\infty = \sqrt{2mE}$.

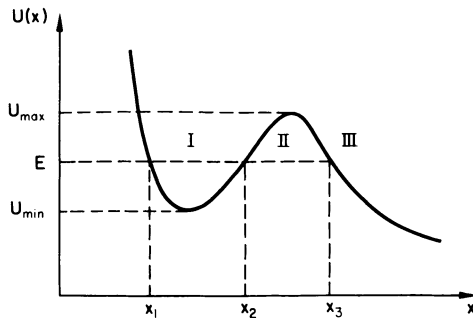


FIG. 9.

If, on the other hand, the particle moves from infinity to the point x_3 , its velocity will gradually decrease and vanish at x_3 , where the particle will turn round and go back to infinity. It cannot penetrate into region I, since this is prevented by the forbidden region II lying between x_2 and x_3 . This region also prevents a particle that is executing oscillations between x_1 and x_2 from entering region III, where motion with energy E is also possible. The forbidden region is called a *potential barrier*, and region I is called a *potential well*. As the particle energy increases in this case, the width of the barrier diminishes and for $E \geq U_{\max}$ it does not exist. The region of oscillatory motion likewise disappears, and the motion of the particle becomes infinite.

Thus we see that the motion of a particle in a given force field may be either finite or infinite depending on the energy of the particle.

This may be illustrated also by the example of motion in a field whose potential-energy curve has the form shown in Fig. 10. In this case positive energies correspond to infinite motion, and negative energies ($U_{\min} < E < 0$) to finite motion.

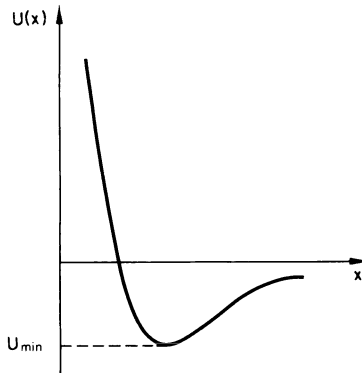


FIG. 10.

Whenever the potential energy is zero at infinity, motion with negative energy will necessarily be finite, since at infinity the zero potential energy exceeds the total energy, and the particle therefore cannot go to infinity.

§14. Elastic collisions

The laws of conservation of energy and momentum can be used to establish relations between various quantities in collisions.

In physics, *collisions* are processes of interaction between bodies in the broad sense of the word, and do not necessarily involve literal contact between the bodies. The colliding bodies are free when at an infinite distance apart. As they pass they interact, and in consequence of this various processes may occur: the bodies may combine, may form new bodies or may undergo an *elastic collision*, in which the bodies move away after their approach, without any change in their internal state. Collisions in which a change occurs in the internal state of the bodies are said to be *inelastic*.

Collisions between ordinary bodies under ordinary conditions are almost always inelastic to some extent, if only because they are accompanied by some heating of the bodies, that is, by the conversion of part of their kinetic energy into heat. Nevertheless, the concept of elastic collisions is of great importance in physics, since such collisions are often involved in physical experiments dealing with atomic phenomena. Ordinary collisions also may frequently be regarded as elastic to a sufficient approximation.

Let us consider an elastic collision between two particles of masses m_1 and m_2 ; let their velocities before and after the collision be respectively $\mathbf{v}_1, \mathbf{v}_2; \mathbf{v}_1', \mathbf{v}_2'$. We shall suppose that the particle m_2 is at rest before the collision, i.e. $\mathbf{v}_2 = 0$.

Since, in an elastic collision, the internal energies of the particles are unchanged, they can be ignored in applying the law of conservation of energy, i.e. they can be taken as zero. Since the particles are assumed not to interact before and after the collision, i.e. to be free, the law of conservation of energy amounts to the conservation of kinetic energy:

$$m_1 v_1^2 = m_1 v_1'^2 + m_2 v_2'^2,$$

where the common factor $\frac{1}{2}$ has been omitted.

The law of conservation of momentum is expressed by the vector equation

$$m_1 \mathbf{v}_1 = m_1 \mathbf{v}_1' + m_2 \mathbf{v}_2'.$$

A very simple case is that where the mass of the particle originally at rest is much greater than that of the incident particle, i.e. $m_2 \gg m_1$. The formula

$$\mathbf{v}_2' = (m_1/m_2)(\mathbf{v}_1 - \mathbf{v}_1')$$

shows that the velocity \mathbf{v}_2' will then be very small. A similar conclusion may be drawn regarding the energy of this particle originally at rest, since the product $m_2 v_2'^2$ will be inversely proportional to the mass m_2 . Hence we deduce that the energy of the first (incident) particle is unchanged by the collision, and its velocity is therefore unchanged in magnitude. Thus a collision between a light and a heavy particle can change only the direction of the velocity of the light particle, the magnitude of its velocity remaining constant.

If the masses of the colliding particles are equal, the conservation laws become

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{v}_1' + \mathbf{v}_2', \\ v_1^2 &= v_1'^2 + v_2'^2. \end{aligned}$$

The first of these relations signifies that the vectors v_1 , v_1' and v_2' form a triangle; the second shows that the triangle is right-angled with hypotenuse v_1 . Thus two particles of equal mass diverge at right angles after the collision (Fig. 11).

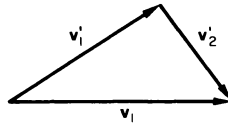


FIG. 11.

Let us next consider a head-on collision of two particles. After such a collision the two particles will move along the direction of the velocity of the incident particle. In this case we can replace the velocity vectors in the law of conservation of momentum by their magnitudes:

$$m_2 v_2' = m_1 (v_1 - v_1').$$

Using also the law of conservation of energy, according to which

$$m_2 v_2'^2 = m_1 (v_1^2 - v_1'^2),$$

we can express v_1' and v_2' in terms of v_1 . Dividing the second expression by the first gives

$$v_2' = v_1 + v_1'$$

and therefore

$$v_1' = \frac{m_1 - m_2}{m_1 + m_2} v_1, \quad v_2' = \frac{2m_1}{m_1 + m_2} v_1.$$

The first (incident) particle will continue to move in the same direction or will move back in the opposite direction, according as its mass m_1 is greater or less than the mass m_2 of the particle originally at rest. If the masses m_1 and m_2 are equal, then $v_1' = 0$, $v_2' = v_1$, so that the particles as if they were exchange velocities. If $m_2 \gg m_1$, then $v_1' = -v_1$ and $v_2' = 0$.

In the general case it is convenient to consider the collision in the centre-of-mass frame of the colliding particles. Then the total

momentum of the particles is zero both before and after the collision. Hence, if the momenta of the first particle before and after the collision are \mathbf{p} and \mathbf{p}' , those of the second particle will be $-\mathbf{p}$ and $-\mathbf{p}'$ respectively.

Next, equating the sums of the kinetic energies of the particles before and after the collision, we see that $p^2 = p'^2$, i.e. the momenta of the particles are unchanged in magnitude. Thus the only effect of the collision is to rotate the momenta of the particles, changing their direction but not their magnitude. The velocities of the two particles are changed in the same manner, being rotated without change of magnitude and remaining opposite in direction, as shown in Fig. 12; the suffix zero to the velocities is used to indicate that they are measured in the centre-of-mass frame.

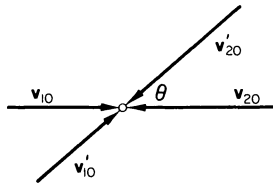


FIG. 12.

The angle through which the velocities are turned is determined not only by the laws of conservation of momentum and energy but also by the nature of the interaction between the particles and by their relative position in a collision.

In order to ascertain how the velocities are changed in the original or *laboratory* frame of reference (in which one particle is at rest before the collision, i.e. $\mathbf{v}_2 = 0$), we use the following graphical procedure. We construct a vector $O1$ equal to the velocity \mathbf{v}_{10} of the first particle in the centre-of-mass frame (Fig. 13). This velocity is related to the velocity \mathbf{v}_1 of the same particle in the laboratory frame (which is also the relative velocity of the two particles) by $\mathbf{v}_{10} = \mathbf{v}_1 - \mathbf{V}$, where

$$\mathbf{V} = \frac{m_1 \mathbf{v}_1 + m_2 \mathbf{v}_2}{m_1 + m_2} = \frac{m_1 \mathbf{v}_1}{m_1 + m_2}$$

is the velocity of the centre of mass. Subtraction gives

$$\mathbf{v}_{10} = \frac{m_2 \mathbf{v}_1}{m_1 + m_2}.$$

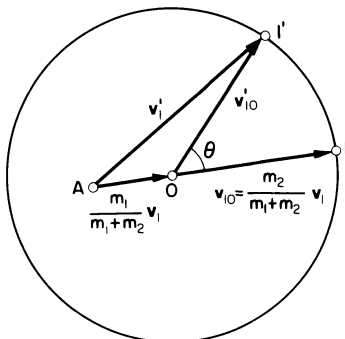


FIG. 13.

The velocity v_{10}' of the first particle after the collision is obtained by turning the velocity v_{10} through some angle θ , i.e. it may be represented by any radius OI' of the circle in Fig. 13. To change to the laboratory frame of reference, we must add to all velocities the velocity V of the centre of mass. In Fig. 13 this is represented by AO . The vector AI then gives the velocity v_1 of the incident particle before the collision, and AI' is the required velocity of that particle after the collision. The velocity of the second particle may be found similarly.

In Fig. 13 it is assumed that $m_1 < m_2$, so that the point A lies within the circle. The vector AI' , i.e. the velocity v_1' , may have any direction. If $m_1 > m_2$, however, A lies outside the circle (Fig. 14). In this case the angle ϕ between the velocities of the particle before and after the collision cannot exceed some maximum value corresponding to AI' being a tangent to the circle. The

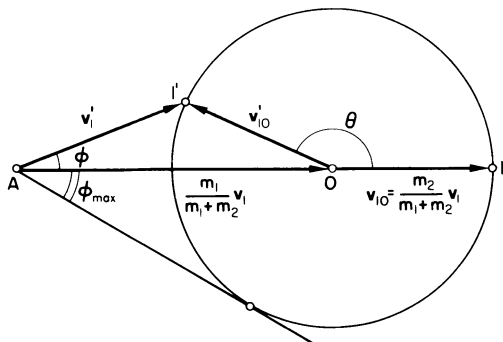


FIG. 14.

side $A1'$ of the triangle $A1'O$ is then perpendicular to $O1'$, and

$$\sin \phi_{\max} = O1'/AO = m_2/m_1.$$

We may also note that the velocity of the particle after the collision cannot be less than a certain minimum value, which is reached when the point $1'$ in Fig. 13 (or Fig. 14) is diametrically opposite to 1. This corresponds to a head-on collision of the particles, and the minimum value of the velocity is

$$v_{1' \min} = \frac{|m_1 - m_2|}{m_1 + m_2} v_1.$$

§15. Angular momentum

Besides energy and momentum, another vector quantity called *angular momentum* is conserved for any closed system. This quantity is the sum of the angular momenta of the individual particles, defined as follows.

Let a particle have momentum \mathbf{p} and let its position relative to some arbitrary origin O be given by the radius vector \mathbf{r} . Then the angular momentum \mathbf{L} of the particle is defined as a vector whose magnitude is

$$L = rp \sin \theta$$

(where θ is the angle between \mathbf{p} and \mathbf{r}) and whose direction is perpendicular to the plane through the directions of \mathbf{p} and \mathbf{r} . The latter condition does not completely define the direction of \mathbf{L} , since it may still be either “up” or “down”. It is customary to define the direction of \mathbf{L} as follows: if a right-handed screw is imagined to turn from the direction of \mathbf{r} towards \mathbf{p} , it will advance in the direction of \mathbf{L} (Fig. 15).

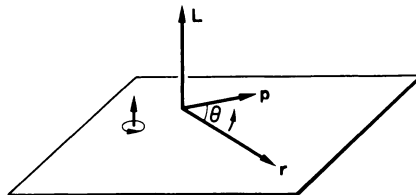


FIG. 15.

The quantity \mathbf{L} may also be regarded in a more intuitive way if we note that the product $r \sin \theta$ is the length h_p of the perpendicular from O to the line of the particle momentum (Fig. 16); this distance is often called the *moment arm* of the momentum relative to O . The angular momentum of the particle is equal to the product of this arm and the magnitude of the momentum:

$$L = ph_p.$$

This vector \mathbf{L} is simply the *vector product* defined in vector algebra; the vector \mathbf{L} constructed in the manner described from the vectors \mathbf{r} and \mathbf{p} is called the vector product of \mathbf{r} and \mathbf{p} and written

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}$$

or, since $\mathbf{p} = m\mathbf{v}$,

$$\mathbf{L} = m\mathbf{r} \times \mathbf{v}.$$

This formula determines the angular momentum of a single particle. The angular momentum of a system of particles is defined as the sum of the individual angular momenta:

$$\mathbf{L} = \mathbf{r}_1 \times \mathbf{p}_1 + \mathbf{r}_2 \times \mathbf{p}_2 + \cdots .$$

This sum is constant in time for any closed system—the *law of conservation of angular momentum*.

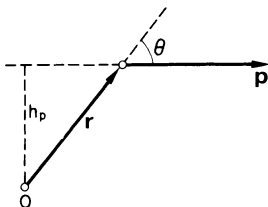


FIG. 16.

It should be noted that the definition of the angular momentum involves an arbitrarily chosen origin O from which the radius vectors of the particles are measured. Although the magnitude

and direction of the vector \mathbf{L} depend on the choice of O , it is easily seen that this dependence does not affect the law of conservation of angular momentum. For if we move the point O through some distance \mathbf{a} of given magnitude and direction, the radius vectors of the particles will all be changed by that amount, and the angular momentum is changed by

$$\mathbf{a} \times \mathbf{p}_1 + \mathbf{a} \times \mathbf{p}_2 + \cdots = \mathbf{a} \times (\mathbf{p}_1 + \mathbf{p}_2 + \cdots) = \mathbf{a} \times \mathbf{P},$$

where \mathbf{P} is the total momentum of the system. For a closed system \mathbf{P} is constant, and we therefore see that changing the origin does not affect the constancy of the total angular momentum of a closed system.

The angular momentum of a system of particles is usually defined with respect to the centre of mass of the system as origin. This will be assumed below.

Let us determine the time derivative of the angular momentum of a particle. The rule for differentiation of a product gives

$$\frac{d\mathbf{L}}{dt} = \frac{d}{dt}(\mathbf{r} \times \mathbf{p}) = \frac{d\mathbf{r}}{dt} \times \mathbf{p} + \mathbf{r} \times \frac{d\mathbf{p}}{dt}$$

Since $d\mathbf{r}/dt$ is the velocity \mathbf{v} of the particle, and $\mathbf{p} = m\mathbf{v}$, the first term is $m\mathbf{v} \times \mathbf{v} = 0$, because the vector product of any vector with itself is zero. In the second term the derivative $d\mathbf{p}/dt$ is, as we know, the force \mathbf{F} acting on the particle. Thus

$$d\mathbf{L}/dt = \mathbf{r} \times \mathbf{F}.$$

The vector product $\mathbf{r} \times \mathbf{F}$ is called the *torque* (relative to a given point O) and will be denoted by \mathbf{K} :

$$\mathbf{K} = \mathbf{r} \times \mathbf{F}.$$

Similarly to the previous discussion of the angular momentum, we can say that the magnitude of the torque is equal to the product of the magnitude F of the force and its moment arm h_F , i.e. the length of the perpendicular from O to the line of action of the force:

$$K = Fh_F.$$

Thus the rate of change of the angular momentum of a particle is equal to the torque acting on it:

$$d\mathbf{L}/dt = \mathbf{K}.$$

The total angular momentum of a closed system is conserved; the time derivative of the sum of the angular momenta of the particles in the system is therefore zero:

$$\frac{d}{dt}(\mathbf{L}_1 + \mathbf{L}_2 + \cdots) = \frac{d\mathbf{L}_1}{dt} + \frac{d\mathbf{L}_2}{dt} + \cdots = 0.$$

Hence it follows that

$$\mathbf{K}_1 + \mathbf{K}_2 + \cdots = 0.$$

We see that in a closed system the sum of the torques is zero, as well as the sum of the forces on all the particles (§7). The latter statement is equivalent to the law of conservation of momentum, and the former to the law of conservation of angular momentum.

There is a profound relation between these properties of a closed system and the fundamental properties of space itself.

Space is *homogeneous*. This means that the properties of a closed system do not depend on its position in space. Let us suppose that a system of particles undergoes an infinitesimal displacement in space, whereby all the particles are moved the same distance in the same direction, and let the vector of this displacement be $d\mathbf{R}$. The work done on the i th particle is $\mathbf{F}_i \cdot d\mathbf{R}$. The sum of the work done must be equal to the change in the potential energy of the system; but since the properties of the system do not depend on its position in space, this change must be zero. Thus we must have

$$\mathbf{F}_1 \cdot d\mathbf{R} + \mathbf{F}_2 \cdot d\mathbf{R} + \cdots = (\mathbf{F}_1 + \mathbf{F}_2 + \cdots) \cdot d\mathbf{R} = 0.$$

Since this equation must hold for any direction of the vector $d\mathbf{R}$, it follows that the sum of the forces $\mathbf{F}_1 + \mathbf{F}_2 + \cdots$ must be zero. We therefore see that the source of the law of conservation of momentum is related to the property of homogeneity of space.

A similar relation exists between the law of conservation of angular momentum and another fundamental property of space, its *isotropy*, i.e. the equivalence of all directions in space. As a result of this isotropy the properties of a closed system are unchanged when the system undergoes any rotation as a whole, and the work done in such a rotation is therefore zero. It can be shown that this condition leads to the vanishing of the sum of the torques in a closed system; we shall return to this topic in §28.

§16. Motion in a central field

The law of conservation of angular momentum is valid for a closed system, and not in general for the individual particles forming the system; but it may in fact be valid for a single particle moving in a force field. For this to be so the field must be a central field.

The term *central field* denotes a force field in which the potential energy of a particle is a function only of its distance r from a certain point, the centre of the field: $U = U(r)$. The force acting on a particle in such a field also depends only on the distance r and is along the radius from the centre to any point in space.

Although a particle moving in such a field is not a closed system, the law of conservation of angular momentum is nevertheless valid for it if the angular momentum is defined relative to the centre of the field. For, since the line of action of the force acting on the particle passes through the centre of the field, the arm of the force about that point is zero, and the torque is therefore zero. From the equation $d\mathbf{L}/dt = \mathbf{K}$ we then have $\mathbf{L} = \text{constant}$.

Since the angular momentum $\mathbf{L} = m\mathbf{r} \times \mathbf{v}$ is perpendicular to the direction of the radius vector \mathbf{r} , the constant direction of \mathbf{L} shows that, as the particle moves, its radius vector must remain in one plane, perpendicular to the direction of \mathbf{L} . Thus particles in a central field move in plane orbits, the plane of each orbit passing through the centre of the field.

The law of conservation of angular momentum in such a “plane” motion may be put in an intuitive form. To do so, we write \mathbf{L} in the form

$$\mathbf{L} = m\mathbf{r} \times \mathbf{v} = m\mathbf{r} \times (ds/dt) = m\mathbf{r} \times ds/dt,$$

where ds is the vector of the displacement of a particle in a time dt . The magnitude of the vector product of two vectors has the geometrical significance of the area of the parallelogram which they form. The area of the parallelogram formed by the vectors ds and r is twice the area of the infinitely narrow sector OAA' (Fig. 17) swept out by the radius vector of the moving particle in time dt . Denoting this area by dS , we can write the magnitude of the angular momentum as

$$L = 2m \, dS/dt.$$

The quantity dS/dt is called the *sectorial velocity*.

Thus the law of conservation of angular momentum can be formulated in terms of the constancy of the sectorial velocity: the radius vector of the moving particle describes equal areas in equal times. In this form it is called *Kepler's second law*.

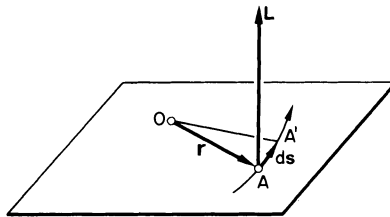


FIG. 17.

The problem of motion in a central field is particularly important because the problem of the relative motion of two interacting particles (the *two-body* problem) can be reduced to it.

Let us consider this motion in the centre-of-mass frame of the two particles. In this frame of reference the total momentum of particles is zero:

$$m_1 \mathbf{v}_1 + m_2 \mathbf{v}_2 = 0,$$

where \mathbf{v}_1 and \mathbf{v}_2 are the velocities of the particles. Let the relative velocity of the particles be

$$\mathbf{v} = \mathbf{v}_1 - \mathbf{v}_2.$$

From these two equations we easily find the formulae

$$\mathbf{v}_1 = \frac{m_2}{m_1 + m_2} \mathbf{v}, \quad \mathbf{v}_2 = -\frac{m_1}{m_1 + m_2} \mathbf{v},$$

which express the velocity of each particle in terms of their relative velocity.

We substitute these formulae in the expression for the total energy of the particles,

$$E = \frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 + U(r),$$

where $U(r)$ is the mutual potential energy of the particles as a function of the distance r between them (i.e. of the magnitude of the vector $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$). A simple reduction then gives

$$E = \frac{1}{2} m v^2 + U(r),$$

where

$$m = m_1 m_2 / (m_1 + m_2)$$

and is called the *reduced mass* of the particles.

We see that the energy of the relative motion of the two particles is the same as if a single particle of mass m were moving with velocity $\mathbf{v} = d\mathbf{r}/dt$ in a central external field with potential energy $U(r)$. Thus the problem of the motion of two particles is equivalent to that of a single “reduced” particle in an external field.

If the solution of the latter problem is known (i.e. if the path $\mathbf{r} = \mathbf{r}(t)$ of the “reduced” particle has been found), we can immediately find the actual paths of the two particles m_1 and m_2 by means of the formulae

$$\mathbf{r}_1 = \frac{m_2}{m_1 + m_2} \mathbf{r}, \quad \mathbf{r}_2 = -\frac{m_1}{m_1 + m_2} \mathbf{r},$$

which express the radius vectors of the particles \mathbf{r}_1 and \mathbf{r}_2 with respect to their centre of mass in terms of their distance apart $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$; these formulae follow from the relation $m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2 = 0$

and correspond to the analogous formulae given above for the velocities $\mathbf{v}_1 = d\mathbf{r}_1/dt$ and $\mathbf{v}_2 = d\mathbf{r}_2/dt$. Hence we see that the two particles will move relative to the centre of mass of the system along geometrically similar paths which differ only in having sizes inversely proportional to the masses of the particles:

$$r_1/r_2 = m_2/m_1.$$

During the motion the particles are always on a line passing through the centre of mass.

CHAPTER II

FIELDS

§17. Electrical interaction

In Chapter I we have given a definition of force and the relation between force and potential energy. We shall now go on to a specific analysis of some of the interactions underlying various physical phenomena.

One of the most important kinds of interaction in Nature is *electrical interaction*. In particular, the forces acting in atoms and molecules are essentially of electrical origin, and this interaction is therefore what mainly determines the internal structure of various bodies.

The forces of electrical interaction depend on the existence of a particular physical characteristic of particles, their *electric charge*. Bodies having no electric charge have no electrical interaction.

If bodies may be regarded as particles, the force of electrical interaction between them is proportional to the product of the charges on the bodies and inversely proportional to the square of the distance between them. This is called *Coulomb's law*. Denoting the electrical interaction force by F , the charges on the bodies by e_1 and e_2 , and the distance between them by r , we can write Coulomb's law in the form

$$F = \text{constant} \times e_1 e_2 / r^2.$$

The force F acts along the line joining the charges, and experiment shows that it is sometimes an attraction, sometimes a repulsion. Charges are therefore said to differ in sign. Bodies having charges of the same sign repel each other, while bodies having charges of opposite signs attract each other. A positive sign of the force in Coulomb's law denotes repulsion, and a negative sign attraction. It does not matter which charges are in fact regarded as positive and which as negative, and the

choice usual in physics is a historical convention. Only a difference in the sign of charges has intrinsic significance. If all negative charges were called positive and *vice versa*, there would be no resulting change in the laws of physics.

Since charges are now introduced for the first time and no units of charge have yet been defined, we can take the proportionality coefficient in Coulomb's law equal to unity: $F = e_1 e_2 / r^2$. This establishes a unit of charge, namely the charge whose force of interaction with another similar charge at a distance of one centimetre is one dyne. This is called the *electrostatic unit of charge*. The system of units based on this choice of the constant coefficient in Coulomb's law is called the *electrostatic* or *CGSE system*. In this system the dimensions of charge are

$$\begin{aligned} [e] &= ([F][r]^2)^{1/2} \\ &= \left(\frac{\text{g. cm}}{\text{sec}^2} \text{cm}^2 \right)^{1/2} = \text{g}^{1/2} \text{cm}^{3/2} \text{sec}^{-1}. \end{aligned}$$

In the SI system of units a larger unit of charge is used, called the *coulomb*:

$$1 \text{ coulomb} = 1 \text{ C} = 3 \times 10^9 \text{ CGSE units of charge.}$$

By means of the expression for the force of electrical interaction we can find the mutual potential energy of two electric charges e_1 and e_2 . If the distance between these charges increases by dr , the work done is $dA = e_1 e_2 dr / r^2$. This is equal to the decrease in the potential energy U . Thus

$$\begin{aligned} -dU &= e_1 e_2 dr / r^2 \\ &= -e_1 e_2 d(1/r), \end{aligned}$$

whence

$$U = e_1 e_2 / r.$$

Strictly speaking, a constant term may also be included here; we have taken it as zero, in order that the potential energy should be zero when the charges are at an infinite distance apart. The potential energy of the interaction of two charges is therefore inversely proportional to the distance between them.

§18. Electric field

Since Coulomb's law involves the product of the charges, the force exerted on a charge e by another charge e_1 can be put in the form

$$\mathbf{F} = e\mathbf{E},$$

where \mathbf{E} is a vector independent of the charge e and determined only by the charge e_1 and the distance r between the charges e and e_1 . This vector is called the *electric field* due to the charge e_1 . Its magnitude is

$$E = e_1/r^2$$

and it is directed along the line joining the positions of the charges e_1 and e . The force on e due to e_1 is thus the product of e and the electric field at e due to e_1 .

Thus we have another way of describing electrical interaction. Instead of saying that particle 1 attracts or repels particle 2, we say that the first particle, whose electric charge is e_1 , creates a particular force field in the surrounding space, namely an electric field; particle 2 does not interact directly with particle 1, but is subject to the field created by the latter.

These two ways of describing the interaction are presented here as being only formally different. In reality, however, this is not so; the concept of the electric field is by no means formal. An analysis of electric (and magnetic) fields which vary with time shows that they can exist in the absence of electric charges and are physically real in the same way as the particles that exist in Nature; however, such problems are outside the scope of the basic ideas concerning interactions of particles that are discussed here in connection with the laws of particle motion.

The electric field created by not one but several electric charges is determined by the following fundamental property of electrical interactions: the electrical interaction between two charges is independent of the presence of a third charge. From this we can conclude that, if there are several charged particles, the electric field which they create is equal to the vector sum of the electric fields produced by each particle separately. In other words, the electric fields created by different charges are simply superposed without affecting one another. This remarkable property of the electric field is called the property of *superposition*.

It should not be thought that the property of superposition of electric fields is a direct consequence of the existence of electrical interaction. In reality, this fundamental property of the electric field is a law of Nature. It applies to other fields besides electric fields and plays a very important part in physics.

Let us apply the property of superposition to determine the electric field of a composite body at large distances from it. If the charges on the particles which compose the body are e_1, e_2, \dots , then the fields which they create at a distance r are

$$E_1 = e_1/r^2, \quad E_2 = e_2/r^2, \dots$$

At large distances from the body we may regard the distances from the various particles as equal and the direction, from the particles to the point considered, as constant. Thus by using the property of superposition to find the total field E due to the body, we can simply take the algebraic sum of the fields E_1, E_2, \dots :

$$E = (e_1 + e_2 + \dots)/r^2.$$

We see that the field of a composite body is the same as the field of a single particle with charge

$$e = e_1 + e_2 + \dots$$

In other words, the charge on the composite body is equal to the sum of the charges on the particles which compose the body and does not depend on their relative position and motion. This is called the *law of conservation of charge*.

In general the electric field is complicated, varying from point to point in both magnitude and direction. To represent it graphically

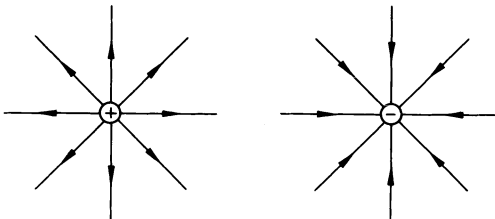


FIG. 18.

we can use *electric lines of force*; these are lines which at every point in space have the direction of the electric field acting at that point.

If the field is created by a single charge, the lines of force are straight lines radiating from the position of the charge, or converging to its position, according as the charge is positive or negative (Fig. 18).

From the definition of the lines of force it is clear that only one line of force passes through each point in space (not occupied by an electric charge), in the direction of the electric field acting at that point. Thus the lines of force do not intersect at points in space where there are no charges.

Electric lines of force in a constant field cannot be closed. For when a charge moves along a line of force, the field forces do a positive amount of work, since the force is always along the path. If there existed closed lines of force, therefore, the work done by the field forces when a charge moved along such a line back to its starting point would not be zero, in conflict with the law of conservation of energy.

Thus the lines of force must necessarily begin and end, or else go to infinity. The points where they begin and end are the charges which create the field. A line of force cannot go to infinity at both ends, since if it did, the field forces would do work when a charge

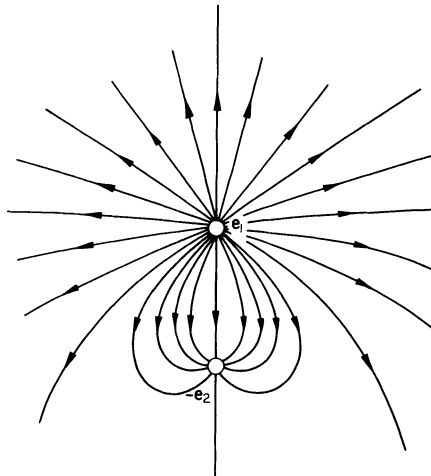


FIG. 19.

is transported along such a line from infinity and back to infinity, in contradiction with the fact that the potential energy is zero at both ends of the path.

One end of a line of force must therefore necessarily be at a charge; the other may go either to infinity or to a charge of the opposite sign. As an illustration, Fig. 19 shows the field of two charges with opposite signs, $+e_1$ and $-e_2$. The diagram is for the case where e_1 is greater than e_2 . Then some of the lines of force leaving $+e_1$ end at the charge $-e_2$, while the others go to infinity.

§19. Electrostatic potential

Like the force, the potential energy U of a charge e in an electric field is proportional to the magnitude of the charge, i.e.

$$U = e\phi.$$

The quantity ϕ which appears here, and which is the potential energy of a unit charge, is called the *potential* of the electric field.

On comparing this definition with that of the field ($\mathbf{F} = e\mathbf{E}$, where \mathbf{F} is the force acting on the charge e) and using the general relation between force and potential energy, $F_s = -dU/ds$ (see §10), we find that a similar relation holds between the field and the potential:

$$E_s = -d\phi/ds.$$

The potential energy of two charges e_1 and e_2 at a distance r is, as we know,

$$U = e_1 e_2 / r.$$

The potential of the field due to a charge e_1 at a distance r is therefore

$$\phi = e_1 / r.$$

With increasing distance from the charge, the potential decreases inversely as the distance.

If the field is due not to one but to several charges e_1, e_2, \dots , it

follows from the principle of superposition that the potential at any point in space is given by the formula

$$\phi = \frac{e_1}{r_1} + \frac{e_2}{r_2} + \dots,$$

where r_i is the distance of the point considered from the charge e_i .

When a charge e moves from a point where the potential is ϕ_1 to a point where it is ϕ_2 , the work done by the field forces is equal to the product of the charge and the difference of potential between the initial and final points:

$$A_{12} = e(\phi_1 - \phi_2).$$

Points at which the potential has a given value lie on a certain surface called an *equipotential* surface. When a charge moves on an equipotential surface, the work done by the field forces is zero. If the work is zero, the force must be perpendicular to the displacement. We can therefore say that the electric field at any point is perpendicular to the equipotential surface through that point. In other words, the lines of force are perpendicular to the equipotential surfaces. For example, for a point charge the lines of force are straight lines passing through the charge, and the equipotential surfaces are concentric spheres with the charge as centre.

The electric potential has dimensions

$$\begin{aligned} [\phi] &= [U]/[e] \\ &= \text{g}^{1/2} \cdot \text{cm}^{1/2} \cdot \text{sec}^{-1}. \end{aligned}$$

This is the unit of potential in the CGSE system. In the SI system a unit 300 times smaller is used, called the *volt*:

$$1 \text{ V} = 1/300 \text{ CGSE unit of potential.}$$

If a charge of one coulomb moves between two points whose potentials differ by one volt, then the work done by the field forces is $3 \times 10^9 \times 1/300 = 10^7$ erg, or one joule:

$$1 \text{ C.V} = 1 \text{ J.}$$

§20. Gauss' theorem

We shall now define the important concept of *electric flux*. To explain this in terms of an analogy, let us imagine the space occupied by an electric field to be filled with some imaginary fluid whose velocity at every point is equal to the electric field in magnitude and direction. The volume of fluid passing through any surface per unit time is equal to the electric flux through that surface.

The electric flux through a spherical surface of radius r due to a point charge e at its centre may be found as follows. The field in this case is, by Coulomb's law, $E = e/r^2$. The velocity of the imaginary fluid is therefore also e/r^2 , and its flux is equal to this velocity multiplied by the area of the sphere, $4\pi r^2$. Thus the flux is

$$E \cdot 4\pi r^2 = 4\pi e.$$

We see that the flux is independent of the radius of the sphere and is determined only by the charge. It may be shown that, if the sphere is replaced by any other closed surface surrounding the charge, the electric flux through it is unchanged and is again equal to $4\pi e$. It should be emphasised that this important result is specifically a consequence of the fact that Coulomb's law involves the inverse square of the distance.

Let us now consider the electric flux due not to one but to several charges. This may be determined by using the superposition property of the electric field. The flux through any closed surface is obviously equal to the sum of the fluxes from the individual charges within that surface. Since each such flux is equal to 4π times the charge, the total electric flux through a closed surface is equal to 4π times the algebraic sum of the charges within the surface. This is called *Gauss' theorem*.

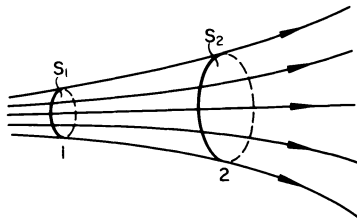


FIG. 20.

If there is no charge within the surface or the total charge within it is zero, the total electric flux through the surface is zero.

Let us consider a narrow bundle of lines of force bounded by a surface itself consisting of lines of force (Fig. 20), and cut this bundle or *tube of force* by two equipotential surfaces 1 and 2; and let us determine the flux through the closed surface formed by the lateral surface of the tube of force and the equipotential surfaces 1 and 2. If there is no charge within this closed surface, the total flux through it will be zero. But the flux through the lateral surface of the tube is obviously zero, and the fluxes through the surfaces 1 and 2 must therefore be equal. The bundle of lines of force may be visualised as a jet of liquid.

Let the fields at the cross-sections 1 and 2 be E_1 and E_2 and the areas of these cross-sections be S_1 and S_2 . Since the tube of force is assumed to be narrow, the fields E_1 and E_2 may be regarded as constant over the respective cross-sections. We can therefore write the equality of the fluxes through the surfaces 1 and 2 as

$$S_1 E_1 = S_2 E_2;$$

since the field is perpendicular to the equipotential surface, the flux is just the product of the field and the surface area. The number N_1 of lines of force passing through the cross-section S_1 is equal to the number N_2 passing through S_2 , and we can therefore write

$$N_1/S_1 E_1 = N_2/S_2 E_2.$$

The quantities $n_1 = N_1/S_1$ and $n_2 = N_2/S_2$ are the numbers of lines of force per unit area of the surfaces 1 and 2, which are orthogonal to the lines of force. Thus we see that the density or concentration of the lines of force is proportional to the field:

$$n_1/n_2 = E_1/E_2.$$

The description of the field by means of the lines of force therefore not only indicates the direction of the field but also gives an idea of its magnitude: where the lines of force are close together the electric field is strong, and where they are far apart it is weak.

§21. Electric fields in simple cases

In many cases Gauss' theorem enables us to find the field due to composite charged bodies if the charge distribution in them is sufficiently symmetrical.

As a first example, let us determine the field of a symmetrically charged sphere. The field of such a sphere is along its radii and depends only on the distance from the centre of the sphere. The field outside the sphere is therefore easily calculated. To do so, let us find the flux through a spherical surface of radius r concentric with the sphere. This flux is evidently $4\pi r^2 E$. By Gauss' theorem, the flux is $4\pi e$, where e is the charge on the sphere. Hence $4\pi r^2 E = 4\pi e$, or

$$E = e/r^2.$$

Thus the field outside the sphere is the same as that of a point charge at the centre of the sphere and equal to the charge on the sphere. Accordingly, the potential is also the same as that of a point charge:

$$\phi = e/r.$$

The field within the sphere depends on how the charges are distributed within the sphere. If all the charges are on the surface of the sphere, then the field within the sphere is zero. If the charge is distributed uniformly through the volume of the sphere with density ρ per unit volume, then the field within the sphere can be found by applying Gauss' theorem to a spherical surface of radius r lying within the sphere:

$$E \cdot 4\pi r^2 = 4\pi e_r,$$

where e_r is the charge within the spherical surface. This charge is equal to the product of the charge density and the volume of a sphere of radius r : $e_r = 4\pi r^3 \rho/3$. Thus

$$4\pi r^2 E = 4\pi \cdot 4\pi r^3 \rho/3,$$

or

$$E = 4\pi \rho r/3.$$

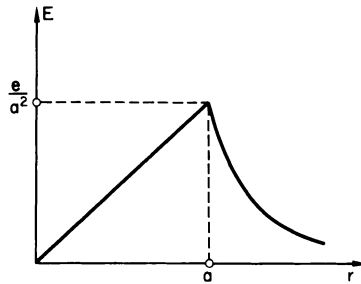


FIG. 21.

We see that the field within a sphere of uniform charge per unit volume is proportional to the distance from the centre, while the field outside the sphere is inversely proportional to the square of this distance. Figure 21 shows the field of such a sphere as a function of the distance from its centre (a denoting the radius of the sphere).

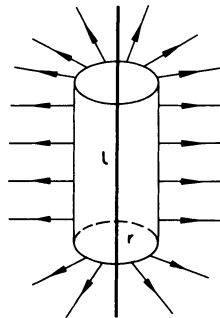


FIG. 22.

As a second example, let us determine the field of a charged straight wire with charges distributed uniformly along it. If the wire is assumed to be sufficiently long, the effect of its ends may be neglected, i.e. it may be regarded as infinitely long. It is evident from symmetry that the field due to such a wire can have no component in either direction along the wire (since the two directions are entirely equivalent), and must therefore be perpendicular to the wire at every point. It is then easy to determine the field of the wire. Let us consider the flux through a closed surface of radius r and length l with its axis along the wire (Fig. 22). Since the field is perpendicular to the axis, the flux

through the ends of the cylinder is zero. The total flux through this closed surface therefore reduces to the flux through the lateral surface of the cylinder, which is evidently $E \cdot 2\pi rl$. By Gauss' theorem, this flux is $4\pi e$, where e is the charge on a length l of the wire; if q denotes the charge per unit length of the wire, then $e = ql$. Thus we have

$$2\pi rlE = 4\pi e = 4\pi ql,$$

whence

$$E = 2q/r.$$

We see that the field due to a uniformly charged wire is inversely proportional to the distance r from the wire.

Let us determine the potential of this field. Since the field \mathbf{E} is along the radius at every point, its radial component E_r is the same as its magnitude E . By the general relation between field and potential we therefore have

$$-d\phi/dr = E = 2q/r,$$

whence

$$\phi = -2q \log_e r + \text{constant}.$$

We see that in this case the potential is a logarithmic function of the distance from the wire. The constant in this formula can not be determined by using the condition that the potential should vanish at infinity, since the above expression becomes infinite as $r \rightarrow \infty$. This is a result of the assumption that the wire is of infinite length, and signifies that the formula derived above can be used only for distances r which are small in comparison with the actual length of the wire.

We may also find the field of a uniformly charged infinite plane. It is evident from symmetry that the field is perpendicular to the plane and has equal values (but opposite directions) at equal distances on either side of the plane.

Let us consider the flux through the closed surface of a rectangular parallelepiped (Fig. 23) bisected by the charged

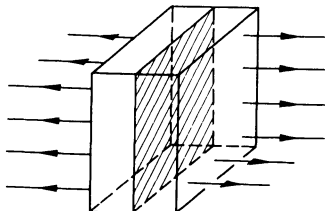


FIG. 23.

plane and having two faces parallel to that plane (the part of the plane lying within the parallelepiped is hatched in the diagram). The only non-zero flux is through these faces. Gauss' theorem therefore gives

$$2SE = 4\pi e = 4\pi S\sigma,$$

where S is the area of the face and σ the charge per unit area of the plane (surface density of charge). Thus we have

$$E = 2\pi\sigma.$$

We see that the field of an infinite plane is independent of the distance from it. In other words, a charged plane creates a uniform electric field on either side of it. The potential of a uniformly charged plane is a linear function of the distance x from it:

$$\phi = -2\pi\sigma x + \text{constant}.$$

§22. Gravitational field

As well as electrical interaction, *gravitational interaction* plays an extremely important part in Nature. This interaction is a property of all bodies, whether they are electrically charged or neutral, and is determined only by the masses of the bodies. The gravitational interaction between all bodies is an attraction, the force of interaction being proportional to the product of the masses of the bodies.

If the bodies may be regarded as particles, the force of gravitational interaction is found to be inversely proportional to the square of the distance between them and proportional to the

product of their masses. Denoting the masses of the bodies by m_1 and m_2 and the distance between them by r , we may write the gravitational force between them as

$$F = -Gm_1m_2/r^2,$$

where G is a universal coefficient of proportionality independent of the nature of the interacting bodies; the minus sign shows that the force F is always attractive. This formula is called *Newton's law of gravitation*.

The quantity G is called the *gravitational constant*; it is evidently the force of attraction between two particles each of unit mass at unit distance apart. The dimensions of the gravitational constant in the CGS system are

$$\begin{aligned} [G] &= [F][r]^2/[m]^2 \\ &= (\text{g.cm.sec}^{-2})\text{cm}^2/\text{g}^2 \\ &= \text{cm}^3/\text{g.sec}^2 \end{aligned}$$

and its value is

$$G = 6.67 \times 10^{-8} \text{ cm}^3/\text{g.sec}^2.$$

The extremely small value of G shows that the force of gravitational interaction can become considerable only for very large masses. For this reason the gravitational interaction plays no part in the mechanics of atoms and molecules. With increasing mass the importance of the gravitational interaction increases, and the motion of bodies such as the Moon, the planets and the artificial satellites is entirely determined by gravitational forces.

The mathematical formulation of Newton's law of gravitation for particles is similar to that of Coulomb's law for point charges. Both the gravitational and the electrical force are inversely proportional to the square of the distance, the mass in the gravitational interaction corresponding to the charge in the electrical interaction. However, whereas electrical forces may be either attractive or repulsive, the gravitational forces are always attractive.

The proportionality coefficient in Coulomb's law has been put equal to unity by appropriate choice of the unit of charge. We

could obviously proceed similarly with Newton's law of gravitation: by putting the gravitational constant equal to unity we should define a certain unit of mass. This would clearly be a derived unit relative to the centimetre and the second, and its dimensions would be cm^3/sec^2 . The new unit of mass would be such as to impart an acceleration of $1 \text{ cm}/\text{sec}^2$ to an equal mass at a distance of 1 cm. Denoting this mass by μ , we can write

$$\begin{aligned} G &= 6.67 \times 10^{-8} \text{ cm}^3/\text{g}.\text{sec}^2 \\ &= 1 \text{ cm}^3/\mu.\text{sec}^2, \end{aligned}$$

whence $\mu = 1.5 \times 10^7 \text{ g} = 15 \text{ tons}$. This new unit is obviously inconvenient, and it is therefore not used, but we can see that in principle a system of units could be constructed in which the only arbitrary units would be those of length and time, and derived units could be constructed for all other quantities, including mass. This system of units is not used in practice, but the possibility of constructing it again shows the arbitrariness of the CGS system.

From the expression for the force of gravitational interaction between two particles we can easily find their potential energy U . Using the general relation between U and F :

$$-dU/dr = F = -Gm_1m_2/r^2,$$

we find

$$U = -Gm_1m_2/r;$$

the arbitrary constant in U is taken as zero so that the potential energy should vanish when the distance between the particles is infinite. This formula is similar to the formula

$$U = e_1e_2/r$$

for the potential energy of the electrical interaction.

We have given above the formulae for the force and potential energy of the gravitational interaction between two particles, but the same formulae are valid for the gravitational forces between any two bodies, provided that the distance between them is large compared with their size. For spherical bodies the

formulae are valid whatever the distance between them, r in this case denoting the distance between the centres of the spheres.

The fact that the gravitational force acting on a particle is proportional to its mass enables us to define the *gravitational field* in the same way as the electric field. The force \mathbf{F} acting on a particle of mass m is written

$$\mathbf{F} = m\mathbf{g},$$

where the field \mathbf{g} depends only on the masses and positions of the bodies which create the field.

Since the gravitational field obeys Newton's law, which is mathematically similar to Coulomb's law for the electric field, Gauss' theorem is valid for the gravitational field also. The only difference is that the charge in Gauss' theorem is now replaced by the mass times the gravitational constant. Thus the gravitational flux through a closed surface is $-4\pi mG$, where m is the total mass within the surface; the minus sign is due to the fact that gravitational forces are attractive.

By using this theorem we can, for example, determine the gravitational field within a uniform sphere. This problem is identical with that of a uniformly charged sphere, discussed in §21. From the result obtained there we can write down immediately

$$g = -4\pi G\rho r/3,$$

where ρ is now the mass density of the sphere.

The gravitational force acting on a body near the Earth's surface is called the *weight* P of the body. The distance of such a body from the centre of the Earth is $R + z$, where R is the Earth's radius and z the altitude of the body above the surface of the Earth. If the altitude z is very small compared with R it may be neglected, and the weight of the body is then

$$P = GmM/R^2,$$

where M is the mass of the Earth. If this formula is written

$$P = mg,$$

then

$$g = GM/R^2.$$

The constant g is then called the *acceleration due to gravity*. It is the acceleration of free fall of a body in the Earth's gravitational field.

At altitudes z such that the force of gravity may be regarded as constant, the potential energy of a body is given by the formula

$$U = Pz = mgz.$$

This can be seen from the general formula derived in §10 for the potential energy in a uniform field, if we use also the fact that in the present case the force is downwards, i.e. in the direction of decreasing z .

In reality, the acceleration due to gravity, g , is not the same at different points on the Earth's surface, since the latter is not perfectly spherical. It should also be remembered that the rotation of the Earth about its axis causes a centrifugal force opposing the force of gravitation. It is therefore necessary to define an effective acceleration due to gravity, which is less than that on a hypothetical non-rotating Earth. At the poles this acceleration is $g = 983.2 \text{ cm/sec}^2$, and at the equator it is $g = 978.0 \text{ cm/sec}^2$.

The value of g sometimes appears in the definition of the units of measurement of physical quantities (e.g. force and work). For this purpose a standard value is arbitrarily defined,

$$g = 980.665 \text{ cm/sec}^2,$$

which is very close to the value at latitude 45° .

§23. The principle of equivalence

The fact that the force of gravity is proportional to the mass of the particle on which it acts ($\mathbf{F} = m\mathbf{g}$) is of very deep physical significance.

Since the acceleration acquired by a particle is equal to the force acting on it divided by the mass, the acceleration \mathbf{w} of a particle in a gravitational field is equal to the field itself,

$$\mathbf{w} = \mathbf{g},$$

and is independent of the mass of the particle. In other words,

the gravitational field has the remarkable property that all bodies, of whatever mass, are equally accelerated by it. This property was first discovered by Galileo in his experiments on the fall of bodies under the Earth's gravity.

A similar behaviour of bodies would be found in a space where no external forces act on the bodies, if their motion were observed in a non-inertial frame of reference. Let us imagine, for example, a rocket in free motion in interstellar space, where the action of gravitational forces may be neglected. Objects within such a rocket will "float", remaining at rest relative to it. If the rocket is given an acceleration \mathbf{w} , however, all the objects in it will "fall" to the floor with acceleration $-\mathbf{w}$. This is the same as would be observed for a rocket moving without acceleration but subject to a uniform gravitational field $-\mathbf{w}$ towards the floor. It would not be possible to distinguish by experiment whether the rocket is moving with an acceleration or is in a uniform gravitational field.

This similarity between the behaviour of bodies in a gravitational field and in a non-inertial frame of reference constitutes what is called the *principle of equivalence*. The fundamental significance of the similarity is fully shown in the theory of gravitation based on the theory of relativity.

In the above discussion we have considered a rocket moving in space in the absence of a gravitational field. The same argument can be "inverted" by considering a rocket moving in a gravitational field, such as that of the Earth. A rocket moving "freely" (i.e. without engines) in such a field will receive an acceleration equal to the field \mathbf{g} . The rocket is then a non-inertial frame of reference, and the effect of this on the motion relative to the rocket of the bodies within it is just balanced by the effect of the gravitational field. This brings about a state of "weightlessness"; that is, objects in the rocket behave as they would in an inertial frame of reference in the absence of any gravitational field. Thus, by considering the motion relative to an appropriately chosen non-inertial frame of reference (in this case, the accelerated rocket), we can as it were "eliminate" the gravitational field. That is, of course, another aspect of the same principle of equivalence.

The gravitational field which "appears" in an accelerated rocket is uniform throughout the rocket and is everywhere equal to $-\mathbf{w}$. Actual gravitational fields, on the other hand, are never uniform. Thus the "elimination" of an actual gravitational field

by changing to a non-inertial frame of reference is possible only within small regions of space, over which the field changes so little that it may be regarded as uniform with sufficient accuracy. In this sense we may say that the gravitational field and the non-inertial frame of reference are only “locally” equivalent.

§24. Keplerian motion

Let us consider the motion of two bodies which attract each other in accordance with the universal law of gravitation, and first suppose that the mass M of one of the bodies is much greater than the mass m of the other body. If the distance r between the bodies is large in comparison with their size, we have a problem of the motion of a particle m in a central gravitational field due to a body M which may be regarded as at rest.

The simplest motion in such a field is uniform motion in a circle round the centre of the field, i.e. round the centre of M . The acceleration is then towards the centre of the circle and is, as we know, equal to v^2/r , where v is the velocity of the particle m . When multiplied by the mass m , this must equal the force exerted on the particle by the body M , i.e.

$$mv^2/r = GmM/r^2,$$

whence

$$v = \sqrt{GM/r}.$$

Using this formula we can, in particular, determine the velocity of an artificial satellite moving near the Earth's surface. Replacing r by the Earth's radius R and GM/R^2 by the acceleration due to gravity g , we obtain as the velocity of the satellite

$$v_1 = \sqrt{GM/R} = \sqrt{gR},$$

called the *first cosmic velocity*. Substituting $g \approx 980 \text{ cm/sec}^2$, $R = 6500 \text{ km}$, we find $v_1 = 8 \text{ km/sec}$.

The above formula for v gives a relation between the radius r of the orbit and the period T of one revolution. Putting

$$v = 2\pi r/T,$$

we find

$$T^2 = 4\pi^2 r^3 / GM.$$

We see that the squares of the periods of revolution are proportional to the cubes of the orbit radii. This is *Kepler's third law*, named after the astronomer who in the early seventeenth century discovered empirically from observations of the planets the fundamental laws of the motion of two bodies under gravitational interaction (called *Keplerian motion*). These laws (of which the second, stating the constancy of the sectorial velocity for motion in a central field, has been discussed in §16) played an important part in Newton's discovery of the universal law of gravitation.

Let us now determine the energy of the particle m . Its potential energy is, as we know,

$$U = -GmM/r.$$

Adding to this the kinetic energy $\frac{1}{2}mv^2$, we find the total energy of the particle:

$$E = \frac{1}{2}mv^2 - GmM/r,$$

which is constant in time.

For motion in a circle we have

$$mv^2 = GmM/r,$$

and therefore

$$E = -\frac{1}{2}mv^2 = -GmM/2r.$$

We see that for motion in a circle the total energy of the particle is negative. This is in agreement with the results of §13, according to which, if the potential energy at infinity is zero, the motion will be finite for $E < 0$ and infinite for $E \geq 0$.

We have discussed a simple circular motion occurring under the action of an attractive force

$$F = -GmM/r^2.$$

In such a field, however, the particle may move not only in a circle but also in an ellipse, hyperbola or parabola. For any of these conic sections one focus (in a parabola, the focus) is at the centre of force (*Kepler's first law*). Elliptical orbits evidently correspond to negative values of the total energy of the particle, $E < 0$ (since the motion is finite). Hyperbolic orbits, with branches which go to infinity, correspond to positive values of the total energy, $E > 0$. Finally, for motion in a parabola $E = 0$. This means that in parabolic motion the velocity of the particle at infinity is zero.

Using the formula for the total energy of the particle, we can easily find the minimum velocity which a satellite must have in order to move in a parabolic orbit, i.e. to escape from the Earth's attraction. Putting $r = R$ in the formula

$$E = \frac{1}{2}mv^2 - GmM/r$$

and equating E to zero, we obtain

$$v_2 = \sqrt{(2GM/R)} = \sqrt{(2gR)},$$

called the *second cosmic velocity* or *velocity of escape*. A comparison with the formula for the first cosmic velocity shows that

$$v_2 = \sqrt{2}v_1 = 11.2 \text{ km/sec.}$$

Let us now see how the parameters of elliptical orbits are defined. The radius of a circular orbit may be expressed in terms of the energy of the particle:

$$r = \alpha/2|E|,$$

where $\alpha = GmM$. When the particle moves in an ellipse, the same formula gives the major semiaxis a of the ellipse:

$$a = \alpha/2|E|.$$

The minor semiaxis b of the ellipse depends not only on the energy but also on the angular momentum L :

$$b = L/\sqrt{(2m|E|)}.$$

The smaller the angular momentum L , the greater the elongation of the ellipse (for a given energy).

The period of revolution in an ellipse depends only on the energy, and is given in terms of the major semiaxis by

$$T^2 = 4\pi^2 ma^3/\alpha.$$

So far we have considered the case where the mass M of one of the bodies is much greater than the mass m of the other body, and we have therefore regarded the body M as being at rest. In reality, of course, both bodies are in motion, and they describe,

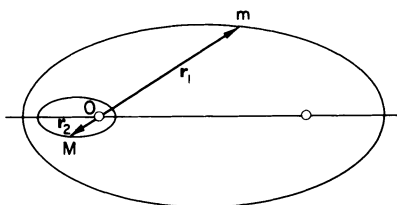


FIG. 24.

in the centre-of-mass frame, geometrically similar paths in the form of conic sections with a common focus at the centre of mass. Fig. 24 shows geometrically similar elliptical orbits of this kind. The particles m and M are at every instant at the ends of a line through the common focus O , and their distances from O are inversely proportional to their masses.

MOTION OF A RIGID BODY

§25. Types of motion of a rigid body

So far we have considered the motion of bodies which might be regarded as particles under certain conditions. Let us now go on to consider motions in which the finite size of bodies is important. Such bodies will be assumed to be *rigid*. In mechanics, this term means that the relative position of the parts of a body remains unchanged during the motion. The body thus moves as a whole.

The simplest motion of a rigid body is one in which it moves parallel to itself; this is called *translation*. For example, if a compass is moved smoothly in a horizontal plane, the needle will retain a steady north-south direction and will execute a translational motion.

In translational motion of a rigid body, every point in it has the same velocity and describes a path of the same shape, there being merely a displacement between the paths.

Another simple type of motion of a rigid body is *rotation* about an axis. In rotation, the various points in the body describe circles in planes perpendicular to the axis of rotation. If in a time dt the body rotates through an angle $d\phi$, the path ds traversed in that time by any point P of the body is clearly $ds = r d\phi$, where r is the distance of P from the axis of rotation. Dividing by dt , we obtain the velocity of P :

$$v = r d\phi/dt.$$

The quantity $d\phi/dt$ is the same at every point of the body and is the angular displacement of the body per unit time. It is called the *angular velocity* of the body, and we shall denote it by Ω .

Thus the velocities at various points in a rigid body rotating about an axis are given by

$$v = r\Omega,$$

where r is the distance of the point from the axis of rotation, the velocity being proportional to this distance.

The quantity Ω in general varies with time. If the rotation is uniform, i.e. the angular velocity is constant, Ω can be determined from the period of rotation T :

$$\Omega = 2\pi/T.$$

A rotation is defined by the direction of the axis of rotation and the magnitude of the angular velocity. These may be combined by means of the angular-velocity vector Ω , whose direction is that of the axis of rotation and whose magnitude is equal to the angular velocity. Of the two directions of the axis of rotation it is customary to assign to the angular-velocity vector the one which is related to the direction of rotation by the "corkscrew rule", i.e. the direction of motion of a right-handed screw rotating with the body.

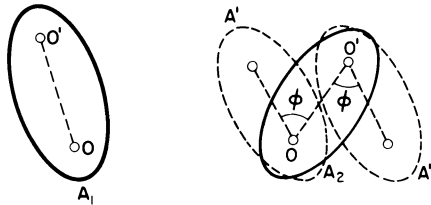


FIG. 25.

These simple forms of motion of a rigid body are especially important because any motion of a rigid body is a combination of translation and rotation. This may be illustrated by the example of a body moving parallel to a certain plane. Let us consider two successive positions of the body, A_1 and A_2 (Fig. 25). The body may evidently be brought from A_1 to A_2 in the following way. We first move the body by a translation from A_1 to a position A' such that some point O of the body reaches its final position. If we then rotate the body about O through a certain angle ϕ , it will reach its final position A_2 .

We see that the complete movement of the body consists of a translation from A_1 to A' and a rotation about O which finally brings the body to the position A_2 . The point O is clearly an arbitrary one: we could equally well carry out a translation of the body from the position A_1 to a position A'' in which some other point O' , instead of O , has its final position, followed by

a rotation about O' which brings the body into its final position A_2 . It is important to note that the angle of this rotation is exactly the same as in the rotation about O , but the distance traversed in the translational motion of the points O and O' is in general different.

The foregoing example shows (what is in fact a general rule) that an arbitrary motion of a rigid body can be represented as a combination of a translational motion of the whole body at the velocity of a point O in it and a rotation about an axis through that point. The translational velocity, which we denote by \mathbf{V} , depends on the point in the body which is chosen, but the angular velocity $\boldsymbol{\Omega}$ does not depend on this choice: whatever the choice of the point O , the axis of rotation passing through it will have the same direction and the angular-velocity magnitude Ω will be the same. In this sense we can say that the angular velocity $\boldsymbol{\Omega}$ is "absolute" and speak of the angular velocity of rotation of a rigid body without specifying the point through which the axis of rotation passes. The translational velocity is not "absolute" in this way.

The "base" point O is usually taken to be the centre of mass of the body. The translational velocity \mathbf{V} is then the velocity of the centre of mass. The advantages of this choice will be explained in §26.

Each of the vectors \mathbf{V} and $\boldsymbol{\Omega}$ is specified by its three components (in some system of coordinates). It is therefore necessary to specify only six independent quantities in order to know the velocity at any point in a rigid body. For this reason a rigid body is said to be a mechanical system with six *degrees of freedom*.

§26. The energy of a rigid body in motion

The kinetic energy of a rigid body in translational motion is very easily found. Since every point in the body is then moving with the same velocity, the kinetic energy is simply

$$E_{\text{kin}} = \frac{1}{2}MV^2,$$

where \mathbf{V} is the velocity of the body and M its total mass. This expression is the same as for a particle of mass M moving with velocity \mathbf{V} . It is clear that translational motion of a rigid body is not essentially different from the motion of a particle.

Let us now determine the kinetic energy of a rotating body. To do so, we imagine it divided into parts so small that they may be regarded as moving like particles. If m_i is the mass of the i th part and r_i its distance from the axis of rotation, then its velocity is $v_i = r_i\Omega$, where Ω is the angular velocity of rotation of the body. The kinetic energy is then $\frac{1}{2}m_iv_i^2$, and summation gives the total kinetic energy of the body:

$$\begin{aligned} E_{\text{kin}} &= \frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2 + \dots \\ &= \frac{1}{2}\Omega^2(m_1r_1^2 + m_2r_2^2 + \dots). \end{aligned}$$

The sum in the parentheses depends on the rigid body concerned (its size, shape and mass distribution) and on the position of the axis of rotation. This quantity characteristic of a given solid body and a given axis of rotation is called the *moment of inertia* of the body about that axis, and is denoted by I :

$$I = m_1r_1^2 + m_2r_2^2 + \dots$$

If the body is continuous it must be divided into an infinite number of infinitesimal parts; the summation in the above formula is then replaced by integration. For example, the moment of inertia of a solid sphere of mass M and radius R , about an axis through its centre, is $2MR^2/5$; that of a thin rod of length l about an axis perpendicular to it through its midpoint is $I = Ml^2/12$.

Thus the kinetic energy of a rotating body may be written as

$$E_{\text{kin}} = \frac{1}{2}I\Omega^2.$$

This expression is formally similar to that for the energy of translation, but the velocity V is replaced by the angular velocity Ω , and the mass by the moment of inertia. This is one example showing that the moment of inertia in rotation corresponds to the mass in translation.

The kinetic energy of a rigid body moving in an arbitrary manner can be written as the sum of the translational and rotational energies if the point O in the method of separating the two motions, described in §25, is taken to be the centre of mass of the body. Then the rotational motion will be a motion of

the points in the body about its centre of mass, and there is an exact analogy with the separation of the motion of a system of particles into the motion of the system as a whole and the “internal” motion of the particles relative to the centre of mass (§12). We saw in §12 that the kinetic energy of the system also falls into two corresponding parts. The “internal” motion is here represented by the rotation of the body about the centre of mass. The kinetic energy of a body moving in an arbitrary manner is therefore

$$E_{\text{kin}} = \frac{1}{2}MV^2 + \frac{1}{2}I_0\Omega^2.$$

The suffix 0 signifies that the moment of inertia is taken about an axis through the centre of mass.

[It should be noted, however, that in this form the result is of practical significance only if the axis of rotation has a constant direction in the body during the motion. Otherwise the moment of inertia has to be taken about different axes at different times, and is therefore no longer a constant.]

Let us consider a rigid body rotating about an axis Z which does not pass through the centre of mass. The kinetic energy of this motion is $E_{\text{kin}} = \frac{1}{2}I\Omega^2$, where I is the moment of inertia about the axis Z . On the other hand, we may regard this motion as consisting of a translational motion with the velocity V of the centre of mass and a rotation (with the same angular velocity Ω) about an axis through the centre of mass parallel to the axis Z . If the distance of the centre of mass from the axis Z is a , then its velocity $V = a\Omega$. The kinetic energy of the body may therefore be written also as

$$\begin{aligned} E_{\text{kin}} &= \frac{1}{2}MV^2 + \frac{1}{2}I_0\Omega^2 \\ &= \frac{1}{2}(Ma^2 + I_0)\Omega^2. \end{aligned}$$

Hence

$$I = I_0 + Ma^2.$$

This formula relates the moment of inertia of the body about any axis to its moment of inertia about a parallel axis through the centre of mass. It is evident that I is always greater than

I_0 . In other words, for a given direction of the axis the minimum value of the moment of inertia is reached when the axis passes through the centre of mass.

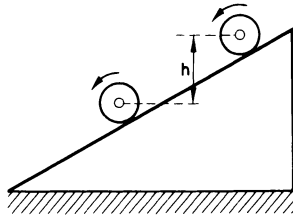


FIG. 26.

If a rigid body moves under gravity, its total energy E is the sum of the kinetic and potential energies. As an example, let us consider the motion of a sphere on an inclined plane (Fig. 26). The potential energy of the sphere is Mgz , where M is the mass of the sphere and z the height of its centre. The law of conservation of energy therefore gives

$$E = \frac{1}{2}MV^2 + \frac{1}{2}I_0\Omega^2 + Mgz = \text{constant}.$$

Let us suppose that the sphere rolls without slipping. Then the velocity v of its point of contact with the plane is zero. On the other hand, this velocity consists of the velocity V of the translational motion of the point down the plane (together with the whole sphere) and the velocity of the point in the opposite direction (up the plane) in its rotation about the centre of the sphere. The latter velocity is ΩR , where R is the radius of the sphere. The equation $v = V - \Omega R = 0$ thus gives

$$\Omega = V/R.$$

Substituting this expression in the law of conservation of energy and assuming that at the initial instant the velocity of the sphere is zero, we find the velocity of the centre of mass of the sphere when it has descended a vertical distance h :

$$V = \sqrt{\left(\frac{2gh}{1 + I_0/MR^2}\right)}.$$

This is, as we should expect, less than the velocity of free fall

of a particle or of a non-rotating body (from the same height h), since the decrease Mgh of the potential energy goes not only to increase the kinetic energy of the translational motion but also to increase that of the rotation of the sphere.

§27. Rotational angular momentum

In rotational motion of a body its angular momentum plays a part similar to that of the momentum in the motion of a particle. In the simple case of a body rotating about a fixed axis Z , this part is played by the angular-momentum component along that axis.

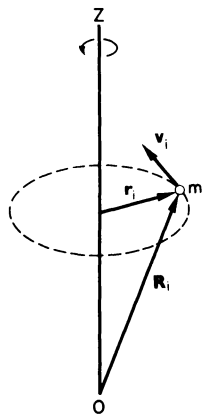


FIG. 27.

To calculate this component, we divide the body into elementary parts, as in calculating the kinetic energy. The angular momentum of the i th element is $m_i \mathbf{R}_i \times \mathbf{v}_i$, where \mathbf{R}_i is the radius vector of this element relative to some point O on the axis Z about which the angular momentum is to be determined (Fig. 27). Since every point in the body moves in a circle round the axis of rotation, the velocity \mathbf{v}_i is tangential to the circle in Fig. 27, i.e. is in a plane at right angles to OZ . We can resolve the vector \mathbf{R}_i into two vectors, one along the axis and the other \mathbf{r}_i perpendicular to the axis. Then the product $m_i \mathbf{r}_i \times \mathbf{v}_i$ is just the part of the angular momentum which is parallel to the axis Z (it will be recalled that the vector product of two vectors is perpendicular to the plane through those vectors). Since the vectors \mathbf{r}_i and \mathbf{v}_i

are mutually perpendicular (being a radius of the circle and a tangent to it), the magnitude of the product $\mathbf{r}_i \times \mathbf{v}_i$ is just $r_i v_i$, where r_i is the distance of the element m_i from the axis of rotation. Finally, since $v_i = \Omega r_i$, we conclude that the component of the angular momentum of the element m_i along the axis of rotation is $m_i r_i^2 \Omega$. The sum $m_1 r_1^2 \Omega + m_2 r_2^2 \Omega + \dots$ gives the required component L_z of the total angular momentum of the body along the axis Z . This quantity is also called the *angular momentum of the body about that axis*.

When the common factor Ω is removed from the above sum there remains a sum which is just the expression for the moment of inertia I . Thus we have finally

$$L_z = I\Omega,$$

i.e. the angular momentum of the body is equal to the angular velocity multiplied by the moment of inertia of the body about the axis of rotation. The analogy between this expression and the expression mv for the momentum of a particle should be noticed: the velocity v is replaced by the angular velocity and the mass is again replaced by the moment of inertia.

If no external forces act on the body, its angular momentum remains constant: it rotates "by inertia" with a constant angular velocity Ω . Here the constancy of Ω follows from that of L_z because we have assumed that the body itself is unchanged during the rotation, i.e. its moment of inertia is unchanged. If the relative position of the parts of the body, and therefore its moment of inertia, vary, then in free rotation the angular velocity will also vary in such a way that the product $I\Omega$ remains constant. For example, if a man holding weights in his hands stands on a platform rotating with little friction, by extending his arms he increases his moment of inertia, and the conservation of the product $I\Omega$ causes his angular velocity of rotation to decrease.

§28. The equation of motion of a rotating body

The equation of motion of a particle gives, as we know, the relation between the rate of change of its momentum and the force acting on it (§7). The translational motion of a rigid body is essentially the same as the motion of a particle, and the equation of this motion consists of the same relation between the total

linear momentum $\mathbf{P} = M\mathbf{V}$ of the body and the total force \mathbf{F} acting on it:

$$d\mathbf{P}/dt = M d\mathbf{V}/dt = \mathbf{F}.$$

In rotational motion there is a corresponding equation relating the rate of change of the angular momentum of the body to the torque acting on it. Let us find the form of this relation, again taking the simple case of rotation of the body about a fixed axis Z .

The angular momentum of the body about the axis of rotation has already been determined. Let us now consider the forces acting on the body. It is clear that forces parallel to the axis of rotation can only move the body along that axis and can not cause it to rotate. We can therefore ignore such forces and consider only those in the plane perpendicular to the axis of rotation.

The corresponding torque K_z about the axis Z is given by the magnitude of the vector product $\mathbf{r} \times \mathbf{F}$, where \mathbf{r} is the vector giving the distance of the point of application of the force \mathbf{F} from the axis. By the definition of the vector product we have

$$K_z = Fr \sin \theta,$$

where θ is the angle between \mathbf{F} and \mathbf{r} ; in Fig. 28 the axis Z is at right angles to the plane of the diagram and passes through O , and A is the point of application of the force. We can also write

$$K_z = h_F F,$$

where $h_F = r \sin \theta$ is the moment arm of the force about the axis (the distance of the line of action of the force from the axis).

According to the relation established in §15 between the rate of change of angular momentum and the applied torque, we can write the equation

$$dL_z/dt = K_z \quad \text{or} \quad I d\Omega/dt = K_z.$$

This is the *equation of motion of a rotating body*. The derivative $d\Omega/dt$ may be called the *angular acceleration*. We see that it is determined by the torque acting on the body, just as the acceleration of the translational motion is determined by the force acting.

If there are several forces acting on the body, then K_z in the above equation must of course be taken as the sum of the torques. It must be remembered that K_z is derived from a vector, and torques tending to turn the body in opposite directions about the axis must be given opposite signs. Those torques have positive signs which tend to rotate the body in the direction in which the angle ϕ of the rotation of the body about the axis is measured (ϕ is the angle whose time derivative is the angular velocity of rotation of the body: $\Omega = d\phi/dt$).

We may also note that the point of application of a force in a rigid body may be displaced in any manner along its line of action without affecting the properties of the motion. This will evidently leave unchanged the arm of the force and therefore the torque.

The condition for equilibrium of a body which can rotate about an axis is evidently that the sum of the torques acting on it should be zero. This is the *law of torques* (or *law of moments*). A particular case is the familiar lever rule which gives the condition of equilibrium for a rod able to rotate about one point in it.

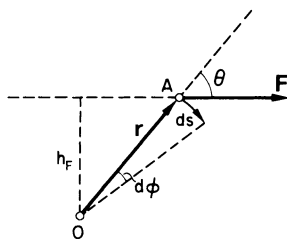


FIG. 28.

There is a simple relation between the torque acting on a body and the work done in a rotation of the body. The work done by a force F when the body rotates about the axis through an infinitesimal angle $d\phi$ (Fig. 28) is equal to the product of the displacement $ds = rd\phi$ of the point A where the force is applied and the component $F_s = F \sin \theta$ of the force in the direction of motion:

$$F_s ds = Fr \sin \theta d\phi = K_z d\phi.$$

We see that the torque about the axis is equal to the work done per unit angular displacement. On the other hand, the work done

on the body is equal to the decrease in its potential energy. We can therefore put $K_z d\phi = -dU$, or

$$K_z = -dU/d\phi.$$

Thus the torque is equal to minus the derivative of the potential energy with respect to the angle of rotation of the body about the given axis. The analogy between this relation and the formula $F = -dU/dx$ should be noticed; the latter relates the force to the change in potential energy in motion of a particle or translational motion of a body.

It is easy to see that the equation of motion of a rotating body is, as it should be, in accordance with the law of conservation of energy. The total energy of the body is

$$E = \frac{1}{2}I\Omega^2 + U,$$

and its conservation is expressed by the equation

$$\frac{d}{dt}(\frac{1}{2}I\Omega^2 + U) = 0.$$

From the rule for differentiating a function of a function we have

$$\frac{dU}{dt} = \frac{dU}{d\phi} \frac{d\phi}{dt} = -K_z \Omega.$$

The derivative $d\Omega^2/dt = 2\Omega d\Omega/dt$. Substituting these expressions and cancelling the common factor Ω , we again obtain the equation $I d\Omega/dt = K_z$.

At the end of §15 it has been mentioned that there is a relation between the law of conservation of the angular momentum of a closed system and the isotropy of space. To establish this relation, we have to prove that the vanishing of the total torque acting in the system is a consequence of the fact that the properties of a closed system are unchanged by any rotation of it as a whole (that is, as if it were a rigid body). By applying the relation $dU/d\phi = -K_z$ to the internal potential energy of the system, taking K_z to be the total torque on all the particles, we see that the condition for the potential energy to be unchanged by a rotation of the closed system about any axis is in fact that the total torque should be zero.

§29. Resultant force

If several forces act on a rigid body, the motion of the body depends only on the total force and the total torque. This sometimes enables us to replace the forces acting on the body by a single force called the *resultant*. It is evident that the magnitude and direction of the resultant are given by the vector sum of the forces, and its point of application must be so chosen that the resultant torque is equal to the sum of the torques.

The most important such case is that of the addition of parallel forces, which includes, in particular, the addition of the forces of gravity acting on the various parts of a rigid body.

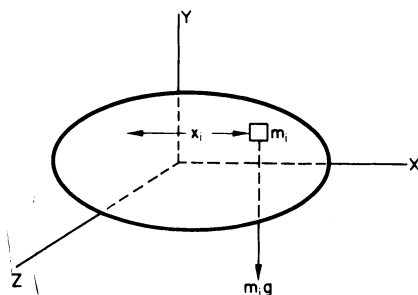


FIG. 29.

Let us consider a body and determine the total torque of gravity about an arbitrary horizontal axis (Z in Fig. 29). The force of gravity acting on the element m_i of the body is $m_i g$, and its moment arm is the coordinate x_i of the element. The total torque is therefore

$$K_Z = m_1 g x_1 + m_2 g x_2 + \dots$$

The magnitude of the resultant is equal to the total weight $(m_1 + m_2 + \dots)g$ of the body, and if the coordinate of its point of application is denoted by X , the torque K_Z has the form

$$K_Z = (m_1 + m_2 + \dots)gX.$$

Equating these two expressions, we find

$$X = (m_1 x_1 + m_2 x_2 + \dots) / (m_1 + m_2 + \dots).$$

This is just the x coordinate of the centre of mass of the body.

Thus we see that all the forces of gravity acting on the body can be replaced by a single force equal to the total weight of the body and acting at its centre of mass. For this reason the centre of mass of the body is often called its *centre of gravity*.

The reduction of a system of parallel forces to a single resultant force is not possible, however, if the sum of the forces is zero. The effect of such forces can be reduced to that of a *couple*, i.e. two forces equal in magnitude and opposite in direction. It is easily seen that the sum K_z of their torques about any axis Z perpendicular to the plane of action of two such forces is equal to the product of either force F and the distance h between their lines of action (the *arm* of the couple):

$$K_z = Fh.$$

The effect of the couple on the motion of the body depends only on this quantity, called the *moment* of the couple.

§30. The gyroscope

In §27 we have derived the component L_z of the angular momentum of a body along the axis of rotation. For a body rotating about a fixed axis, only this component of the vector \mathbf{L} is important. The simple relation between this component and the angular velocity of rotation $\boldsymbol{\Omega}$ ($L_z = I\Omega$) has the result that the entire motion is simple.

If the axis of rotation is not fixed, however, it is necessary to consider the entire vector \mathbf{L} as a function of the angular-velocity vector $\boldsymbol{\Omega}$. This function is more complicated: the components of the vector \mathbf{L} are linear functions of those of $\boldsymbol{\Omega}$, but the directions of the two vectors are in general different. This considerably complicates the nature of the motion of the body in the general case.

Here we shall consider only one example of the motion of a body with a freely variable axis of rotation, namely the *gyroscope*; this is an axially symmetric body rotating rapidly about its geometrical axis.

In such a rotation the angular momentum \mathbf{L} is along the axis of the body, like the angular-velocity vector $\boldsymbol{\Omega}$. This is obvious simply from considerations of symmetry: since the motion is

axially symmetric, there is no other preferred direction which could be taken by the vector \mathbf{L} .

So long as no external forces act on the gyroscope, its axis will remain in a fixed direction in space, since by the law of conservation of angular momentum the direction (and the magnitude) of the vector \mathbf{L} does not vary. If external forces are applied to the gyroscope, its axis will begin to deviate. It is this movement of the gyroscope axis (called *precession*) which we shall discuss.

The change in direction of the gyroscope axis consists in a rotation about some other axis, so that the total angular-velocity vector is not along the geometrical axis of the body. The angular-momentum vector \mathbf{L} likewise will not coincide with this axis (nor with the direction of $\boldsymbol{\Omega}$). But if the primary rotation of the gyroscope is sufficiently rapid, and the external forces are not too great, the rate of rotation of the gyroscope axis will be relatively small, and the vector $\boldsymbol{\Omega}$ (and therefore \mathbf{L}) will always be in a direction close to the axis. Hence, if we know how the vector \mathbf{L} varies, we know approximately how the gyroscope axis moves. The change in angular momentum is given by the equation

$$d\mathbf{L}/dt = \mathbf{K},$$

where \mathbf{K} is the applied torque.

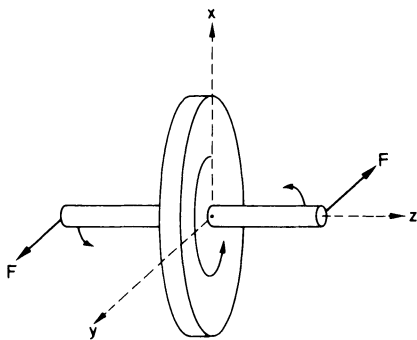


FIG. 30.

For example, let forces F acting in the yz plane be applied at the ends of the gyroscope axis (the z axis in Fig. 30). Then the moment \mathbf{K} of the couple is along the x axis, and the derivative

$d\mathbf{L}/dt$ will also be in that direction. Hence the angular momentum \mathbf{L} , and therefore the gyroscope axis, will deviate in the direction of the x axis.

Thus the application of a force to the gyroscope causes its axis to turn in a direction perpendicular to the force.

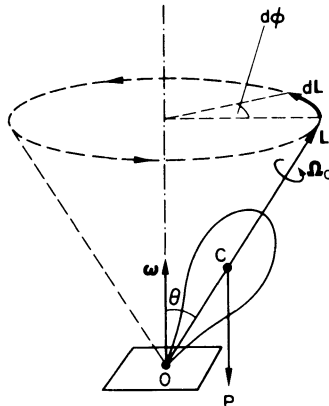


FIG. 31.

An example of a gyroscope is a top supported at its lowest point. [In the following discussion we neglect friction at the support.] The top is subject to the force of gravity, whose direction is constant, namely vertically downwards. This force is equal to the weight of the top: $P = Mg$, where M is the mass of the top, and acts at its centre of gravity (C in Fig. 31). The torque about the point of support O is in magnitude $K = Pl \sin \theta$ (where l is the distance OC and θ the angle between the axis of the top and the vertical), and its direction is always perpendicular to the vertical plane through the axis of the top. Under the action of this torque the vector \mathbf{L} (and therefore the axis of the top) will be deflected, remaining constant in magnitude and at a constant angle θ to the vertical, i.e. describing a cone about the vertical.

It is easy to determine the angular velocity of precession of the top. We denote this by $\boldsymbol{\omega}$ to distinguish it from that of the rotation of the top about its own axis, which is denoted by $\boldsymbol{\Omega}_0$.

In an infinitesimal time dt the vector \mathbf{L} receives an increment $d\mathbf{L} = \mathbf{K}dt$ perpendicular to itself and lying in a horizontal plane (Fig. 31).

Dividing this by the component of the vector \mathbf{L} in this plane, we find the angle $d\phi$ through which the component turns in the time dt :

$$d\phi = \frac{K}{L \sin \theta} dt.$$

The derivative $d\phi/dt$ is evidently the required angular velocity of precession. Thus

$$\omega = K/L \sin \theta.$$

Substituting $K = Mgl \sin \theta$ and $L = I\Omega_0$ (where I is the moment of inertia of the top about its axis), we obtain finally

$$\omega = Mgl/I\Omega_0.$$

The rotation of the top, it will be remembered, has been assumed sufficiently rapid. We can now make this condition more precise: we must have $\Omega_0 \gg \omega$.

Since

$$\omega/\Omega_0 = Mgl/I\Omega_0^2,$$

we see that this condition implies that the potential energy of the top in the gravitational field ($Mgl \cos \theta$) must be small in comparison with its kinetic energy ($\frac{1}{2}I\Omega_0^2$).

§31. Inertia forces

So far we have considered the motion of bodies with respect to inertial frames of reference, and have discussed only in §23 a frame of reference in accelerated translational motion (an accelerated rocket). We have seen that, from the point of view of an observer moving with the rocket, the fact that the frame of reference is non-inertial is perceived through the appearance of a force field equivalent to a uniform field of gravity.

The additional forces which appear in non-inertial frames of references are called *inertia forces*. Their characteristic feature is that they are proportional to the mass of the body on which they act. This makes them similar to gravitational forces.

Let us now consider how motion occurs with respect to a rotating frame of reference, and the inertia forces which appear. The Earth itself, for example, is such a frame of reference; owing to the Earth's daily rotation, the frame of reference in which the Earth is fixed is, strictly speaking, non-inertial, although the resulting inertia forces are comparatively small because of the slowness of the rotation.

For simplicity, let us assume that the frame of reference is a disc rotating uniformly (with angular velocity Ω) and consider a simple motion on it: that of a particle moving uniformly along the edge of the disc. Let the velocity of this particle relative to the disc be v_n , the suffix n indicating that the frame of reference is non-inertial. The velocity v_i of the particle relative to a fixed observer (inertial frame of reference) is evidently the sum of v_n and the velocity of the points on the edge of the disc itself. The latter is Ωr , where r is the radius of the disc. Hence

$$v_i = v_n + \Omega r.$$

It is easy to determine the acceleration w_i of the particle in the inertial frame of reference. Since the particle moves uniformly in a circle of radius r with velocity v_i , we have

$$\begin{aligned} w_i &= v_i^2/r \\ &= v_n^2/r + 2\Omega v_n + \Omega^2 r. \end{aligned}$$

Multiplying this acceleration by the mass m of the particle, we find the force F acting on the particle in the inertial frame of reference:

$$F = mw_i.$$

Let us now consider how this motion will appear to an observer located on the disc and regarding it as being at rest. This observer also will see the particle moving uniformly in a circle of radius r , but with velocity v_n . The acceleration of the particle relative to the disc is therefore

$$w_n = v_n^2/r$$

towards the centre of the disc. Regarding the disc as being at rest, the observer multiplies w_n by the mass of the particle and takes this product to be the force F_n acting on the particle:

$$F_n = mw_n.$$

Since

$$w_n = w_i - 2\Omega v_n - \Omega^2 r,$$

and $mw_i = F$, we find that

$$F_n = F - 2m\Omega v_n - m\Omega^2 r.$$

Thus we see that, relative to the rotating frame of reference, the particle is subject not only to the “true” force F but also to two additional forces, $-m\Omega^2 r$ and $-2m\Omega v_n$. The former of these inertia forces is called the *centrifugal force*, and the latter the *Coriolis force*. The minus signs indicate that in this case both forces are directed away from the axis of rotation of the disc.

The centrifugal force is independent of the velocity v_n , i.e. it exists even if the particle is at rest relative to the disc. For a particle at a distance r from the axis of rotation of the frame of reference this force is always equal to $m\Omega^2 r$ and is directed radially away from the axis.

Having defined the centrifugal force, we may also define the *centrifugal energy* as the potential energy of a particle in the centrifugal force field. According to the general formula relating the force and the potential energy, we have

$$-dU_{cf}/dr = m\Omega^2 r,$$

whence

$$U_{cf} = -\frac{1}{2}m\Omega^2 r^2 + \text{constant}.$$

The arbitrary constant may reasonably be taken as zero, the potential energy (thus being measured from its value on the axis of rotation ($r = 0$), where the centrifugal force is zero.

The centrifugal force can reach very large values in specially

designed centrifuges. On the Earth, it is very small. Its maximum value occurs at the equator, where the force on a particle of mass 1 g is

$$\begin{aligned} m\Omega^2 R &= 1 \times (2\pi/24 \times 60 \times 60)^2 \times 6.3 \times 10^8 \text{ dyn} \\ &= 3.3 \text{ dyn} \end{aligned}$$

($R = 6.3 \times 10^8$ cm being the Earth's radius). This force therefore decreases the weight of a body by 3.3 dyne per gram, i.e. by about 0.3%.

The second inertia force, the Coriolis force, is quite different in type from any of the forces so far discussed. It acts only on a particle which is in motion (relative to the frame of reference considered) and depends on the velocity of that motion. It is, on the other hand, independent of the position of the particle relative to the frame of reference. In the example discussed above, its magnitude is $2m\Omega v_n$ and its direction is away from the axis of rotation of the disc. It can be shown that in general the Coriolis inertia force on a particle moving with any velocity \mathbf{v}_n relative to a frame of reference rotating with angular velocity $\boldsymbol{\Omega}$ is

$$2m\mathbf{v}_n \times \boldsymbol{\Omega}.$$

In other words, it is perpendicular to the axis of rotation and to the velocity of the particle, and its magnitude is $2mv_n\Omega \sin \theta$, where θ is the angle between \mathbf{v}_n and $\boldsymbol{\Omega}$. When the direction of the velocity \mathbf{v}_n is reversed, so is that of the Coriolis force.

Since the Coriolis force is always perpendicular to the direction of motion of the particle, it does no work, but simply changes the direction of motion of the particle without altering its velocity.

Although the Coriolis force is usually very small on the Earth, it does bring about certain specific effects. Because of this force, a freely falling body will not move exactly vertically, but will be deflected slightly eastwards. The deviation is very slight, however. For example, calculation shows that the deflection in a fall from a height of 100 m in latitude 60° is only about 1 cm.

The Coriolis force accounts for the behaviour of the Foucault pendulum, which used to serve as a demonstration of the Earth's rotation. If there were no Coriolis force, the plane of oscillation of a pendulum swinging near the Earth's surface would remain

fixed (relative to the Earth). The effect of this force is to cause the plane of oscillation to rotate round the vertical with angular velocity $\Omega \sin \phi$, where Ω is the angular velocity of the Earth's rotation and ϕ the latitude of the point at which the pendulum is suspended.

The Coriolis force plays a large part in meteorological phenomena. For example, the trade winds, which blow from the tropics to the equator, would blow directly from the north in the northern hemisphere and from the south in the southern hemisphere, if the Earth were not rotating. The Coriolis force causes a westward deflection of these winds.

CHAPTER IV
OSCILLATIONS

§32. **Simple harmonic oscillations**

We have seen in §13 that a one-dimensional motion of a particle in a potential well is periodic, i.e. is repeated at equal intervals of time. The interval after which the motion is repeated is called the *period* of the motion. If this is denoted by T , then the particle has the same position and velocity at times t and $t + T$.

The reciprocal of the period is called the *frequency*, and will be denoted by ν :

$$\nu = 1/T;$$

it gives the number of times per second that the motion is repeated. Its dimensions are evidently 1/sec, and the unit of measurement of frequency, corresponding to a period of 1 sec, is called the *hertz* (Hz): $1 \text{ Hz} = 1 \text{ sec}^{-1}$.

There is obviously an infinite variety of types of periodic motion. The simplest periodic functions are the trigonometric sine and cosine functions, and the simplest periodic motion is therefore one in which the coordinate of the particle varies according to

$$x = A \cos (\omega t + \alpha),$$

where A , ω and α are constants. Such a periodic motion is called a *simple harmonic oscillation*.

The quantities A and ω have a simple physical significance. Since the period of the cosine is 2π , the period T of the motion is related to ω by

$$T = 2\pi/\omega.$$

Hence we see that ω differs from ν by a factor 2π :

$$\omega = 2\pi\nu.$$

The quantity ω is called the *angular frequency*; it is generally used in physics to describe oscillations, and is often called simply the *frequency*.

Since the maximum value of the cosine is unity, the maximum value of the coordinate x is A . This maximum value is called the *amplitude* of the oscillation, and x varies from $-A$ to A .

The argument $\omega t + \alpha$ of the cosine is called the *phase* of the oscillation, and α is the *initial phase* (at time $t = 0$).

The velocity of the particle is

$$v = dx/dt = -A\omega \sin(\omega t + \alpha).$$

We see that the velocity also varies harmonically but the cosine is replaced by a sine. If this expression is written

$$v = A\omega \cos(\omega t + \alpha + \frac{1}{2}\pi)$$

we can say that the velocity “leads” the coordinate by $\frac{1}{2}\pi$ in phase. The amplitude of the velocity is equal to the amplitude of the displacement multiplied by the frequency ω .

Let us now ascertain what the force acting on the particle must be in order to cause it to execute simple harmonic oscillations. To do so, we find the acceleration of the particle in such a motion. We have

$$w = dv/dt = -A\omega^2 \cos(\omega t + \alpha).$$

This quantity varies in the same manner as the coordinate of the particle, but differs from it in phase by π . Multiplying w by the mass m of the particle and noticing that $A \cos(\omega t + \alpha) = x$, we obtain the following expression for the force:

$$F = -m\omega^2 x.$$

Thus, in order that a particle should execute simple harmonic oscillations, the force acting on it must be proportional to the

displacement of the particle and in the opposite direction. An elementary example is that of the force exerted on a body by a stretched (or compressed) spring; this is proportional to the elongation (or shortening) of the spring and is always in a direction such that the spring tends to regain its original length. Such a force is often called a *restoring force*.

The dependence of the force on the position of a particle in physical problems is very often found to be of this type. If a body is in a position of stable equilibrium (at the point $x = 0$, say) and is then moved slightly in either direction from this position, a force F results which tends to return the body to its equilibrium position. As a function of the position x of the body, the force $F = F(x)$ is represented by a curve passing through the origin: at the point $x = 0$ the force $F = 0$, and it has opposite signs on either side of this point. Over a short range of values of x , this curve can be approximated by a section of a straight line, so that the force is proportional to the displacement x . Thus, if the body undergoes a slight displacement from the equilibrium position and is then left to itself, its return to the equilibrium position will give rise to simple harmonic oscillations.

Motions in which a body deviates only slightly from a position of equilibrium are called *small oscillations*. Thus small oscillations are simple harmonic. The frequency of these oscillations is determined by the rigidity with which the body is fixed; this gives the relation between the force and the displacement. If the force is related to the displacement by

$$F = -kx,$$

where k is a coefficient called the *stiffness*, a comparison with the expression for the force in simple harmonic motion, $F = -m\omega^2x$, shows that the frequency of the oscillations is

$$\omega = \sqrt{k/m}.$$

It must be emphasised that the frequency depends only on the properties of the oscillating system (the rigidity with which the body is fixed, and the mass of the body), and not on the amplitude of the oscillations. A given body executing oscillations of various amplitudes does so with the same frequency. This is a very

important property of small oscillations. The amplitude, on the other hand, is determined not by the properties of the system itself but by the initial conditions of its motion, i.e. by the initial disturbance which causes the system to be no longer at rest. The oscillations of the system resulting from an initial disturbance, after which the system is left to itself, are called *natural oscillations*.

The potential energy of an oscillating particle is easily found by noting that

$$dU/dx = -F = kx,$$

whence

$$U = \frac{1}{2}kx^2 + \text{constant.}$$

Choosing the constant so that the potential energy is zero in the equilibrium position ($x = 0$), we have finally

$$U = \frac{1}{2}kx^2,$$

i.e. the potential energy is proportional to the square of the displacement of the particle.

Adding the potential energy to the kinetic energy, we find the total energy of the oscillating particle:

$$\begin{aligned} E &= \frac{1}{2}mv^2 + \frac{1}{2}kx^2 \\ &= \frac{1}{2}mA^2\omega^2 \sin^2(\omega t + \alpha) + \frac{1}{2}mA^2\omega^2 \cos^2(\omega t + \alpha) \end{aligned}$$

or

$$E = \frac{1}{2}mA^2\omega^2.$$

Thus the total energy is proportional to the square of the amplitude of the oscillations. It should be noted that the kinetic and potential energies vary as $\sin^2(\omega t + \alpha)$ and $\cos^2(\omega t + \alpha)$, so that when one increases the other decreases. In other words, the process of oscillation involves a periodic transfer of energy between potential and kinetic and *vice versa*. The mean values (over the period of the oscillation) of the potential and kinetic energies are each equal to $\frac{1}{2}E$.

§33. The pendulum

As an example of small oscillations, let us consider oscillations of a *simple pendulum*; this consists of a particle suspended by a string in the Earth's gravitational field.

Let us deflect the pendulum from its equilibrium position through an angle ϕ and determine the force then acting on it. The total force on the pendulum is mg , where m is the mass of the pendulum and g the acceleration due to gravity. We resolve this force into two components (Fig. 32), one along the string and the other perpendicular to it. The first component is balanced by the tension in the string, while the second component causes the motion of the pendulum. This component is evidently

$$F = -mg \sin \phi.$$

For small oscillations the angle ϕ is small, and $\sin \phi$ is approximately equal to ϕ itself, so that $F \approx -mg\phi$. Since $l\phi$ is the distance x through which the particle moves (l being the length of the pendulum), we can write

$$F = -mgx/l.$$

Thus we see that the stiffness $k = mg/l$ for small oscillations of a pendulum. The frequency of these oscillations is therefore

$$\omega = \sqrt{(g/l)}.$$

The period of the oscillations is

$$T = 2\pi/\omega = 2\pi\sqrt{(l/g)}.$$

The length of a pendulum with period $T = 1$ sec, for the standard acceleration due to gravity given at the end of §22, is $l = 24.84$ cm.

The manner in which the period of a pendulum depends on its length and the acceleration due to gravity can also be easily determined from dimensional considerations. The quantities available to characterise the mechanical system in question are m , l and g , with dimensions

$$[m] = \text{g}, \quad [l] = \text{cm}, \quad [g] = \text{cm/sec}^2.$$

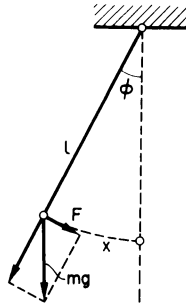


FIG. 32.

The period T can depend only on these quantities. Since only m contains the dimension g , and the dimensions of the required quantity $[T] = \text{sec}$ do not contain g , it is clear that T cannot depend on m . From the two remaining quantities l and g we can eliminate the dimension cm (which is not present in T) by taking the ratio l/g . Finally, by taking the square root $\sqrt{l/g}$ we obtain a quantity having the dimensions sec , and it is clear from the foregoing argument that this is the only way in which such a quantity can be obtained. We can therefore assert that the period T must be proportional to $\sqrt{l/g}$; the numerical value of the coefficient of proportionality can not, of course, be determined by this method.

So far we have discussed small oscillations in terms of a single particle, but the results obtained in fact apply also to the oscillations of more complex systems. As an example, let us consider the oscillations of a rigid body that can rotate under gravity about a horizontal axis. This is called a *compound pendulum*.

We have seen in §28 that the laws of motion of a rotating body are formally identical with those of a particle, the coordinate x being replaced by the angle of rotation ϕ , the mass by the moment of inertia I of the body about the axis of rotation, and the force F by the torque K_z .

In the present case the torque of gravity about the axis of rotation is $K_z = -mga \sin \phi$, where m is the mass of the body, a the distance of its centre of gravity C from the axis of rotation (which passes through the point O at right angles to the plane of Fig. 33), and ϕ the angle of deflection of the line OC from the vertical. The minus sign shows that the torque K_z tends to

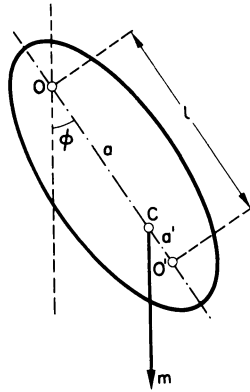


FIG. 33.

decrease the angle ϕ . In small oscillations the angle ϕ is small, and therefore $K_z \approx -mga\phi$. Comparing this with the expression for the restoring force $F = -kx$ in oscillations of a particle, we see that the stiffness coefficient k is now replaced by mga . Thus, by analogy with the formula $\omega = \sqrt{(k/m)}$, we can write down the following expression for the frequency of oscillations of a compound pendulum:

$$\omega = \sqrt{(mga/I)}.$$

A comparison of this with the formula $\omega = \sqrt{(g/l)}$ for the frequency of oscillations of a simple pendulum shows that the properties of the motion of a compound pendulum are the same as those of a *simple equivalent pendulum* of length

$$l = I/ma.$$

Putting $I = I_0 + ma^2$ (where I_0 is the moment of inertia of the pendulum about a horizontal axis through the centre of gravity), we can write the equivalent length as

$$l = a + I_0/ma.$$

From this expression we can draw the following interesting conclusion. If we mark off $OO' = l$ along the line OC (Fig. 33), and now suppose that the pendulum is suspended from an axis

passing through O' , then the equivalent length of the resulting pendulum is

$$l' = a' + I_0/ma'.$$

But $a' = l - a = I_0/ma$, and therefore $l' = l$. Thus the equivalent lengths, and therefore the periods of oscillation, are the same for pendulums suspended from axes at a distance l apart.

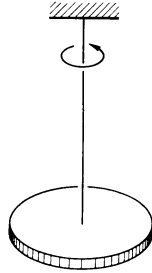


FIG. 34.

Finally, let us consider *torsional oscillations* of a disc suspended on an elastic wire (Fig. 34). The elastic torque which is created when the wire is twisted and which tends to restore the disc to its original position is proportional to the angle ϕ of rotation of the disc: $K_z = -k\phi$, where k is a constant coefficient depending on the properties of the wire. If the moment of inertia of the disc about its centre is I_0 , the frequency of oscillations is

$$\omega = \sqrt{(k/I_0)}.$$

§34. Damped oscillations

So far we have considered the movement (including oscillations) of bodies as if they occurred completely without hindrance. If a motion takes place in an external medium, however, the latter will resist the motion and tend to retard it. The interaction between a body and a medium is a complicated process which ultimately causes the energy of the moving body to be transformed into heat—the *dissipation of energy*, as it is called in physics. This process is not a purely mechanical one, and a detailed study of it involves other branches of physics also.

From a purely mechanical point of view it can be described by defining a certain additional force which appears as a result of the motion itself and is in the opposite direction to the motion. This force is called *friction*. For sufficiently small velocities it is proportional to the velocity of the body:

$$F_{\text{fr}} = -bv,$$

where b is a positive constant describing the interaction between the body and the medium, and the minus sign indicates that the force is in the opposite direction to the velocity.

Let us see what is the effect of such friction on an oscillatory motion. We shall suppose that the friction is so small that the resulting energy loss by the body in one period of oscillation is relatively small.

The energy loss is defined as the work done by the friction. In a time dt this work, and therefore the energy loss dE , is equal to the product of the force F_{fr} and the displacement $x = v dt$ of the body:

$$dE = F_{\text{fr}} dx = -bv^2 dt,$$

whence

$$dE/dt = -bv^2 = -\frac{2b}{m} \cdot \frac{1}{2}mv^2.$$

On the above assumption that the friction is small, we can apply this formula to the mean energy loss over one period, replacing the kinetic energy $\frac{1}{2}mv^2$ by its mean value also. We have seen in §32 that the mean value of the kinetic energy of an oscillating body is half its total energy E . Thus we can write

$$dE/dt = -2\gamma E,$$

where $\gamma = b/2m$. We see that the rate of decrease of the energy is proportional to the energy itself.

Writing this relation in the form

$$dE/E = d(\log_e E) = -2\gamma dt,$$

we find $\log_e E = -2\gamma t + \text{constant}$, or finally

$$E = E_0 e^{-2\gamma t},$$

where E_0 is the value of the energy at the initial instant ($t = 0$).

Thus the energy of the oscillations is reduced exponentially by friction. The amplitude A of the oscillations decreases with the energy; since the energy is proportional to the square of the amplitude, we have

$$A = A_0 e^{-\gamma t}.$$

The decrease of the amplitude is determined by the *damping coefficient* γ . In a time $\tau = 1/\gamma$ the amplitude decreases by a factor e ; this is called the *time constant* of the decay of the oscillations. Our hypothesis that the friction is slight means that τ is assumed large in comparison with the period $T = 2\pi/\omega$ of the oscillations. The small quantity T/τ is called the *logarithmic damping decrement*.

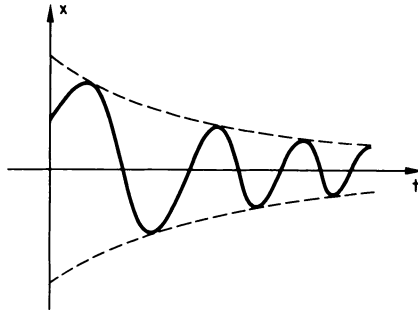


FIG. 35.

Figure 35 shows a graph of the displacement as a function of time for the damped oscillations

$$\begin{aligned} x &= A \cos(\omega t + \alpha) \\ &= A_0 e^{-\gamma t} \cos(\omega t + \alpha). \end{aligned}$$

The broken lines show the decrease of the amplitude.

Friction also affects the frequency of the oscillations. By retarding the motion it increases the period, i.e. decreases the frequency of the oscillations. When the friction is slight, however, this change is very small, and has therefore been ignored above; it may be shown that the relative change in frequency is proportional to the square of the small quantity γ/ω . When the friction is sufficiently great, however, the retardation may be so considerable that the motion is damped without oscillation; this is called *aperiodic damping*.

§35. Forced oscillations

In any actual oscillating system, friction processes of some kind always occur. The natural oscillations which result from the action of an initial disturbance are therefore damped in the course of time.

In order to produce undamped oscillations in a system it is necessary to compensate the energy losses due to friction. This may be done by sources of energy external to the oscillating system. A simple case is the action on the system of a variable external force F_{ext} which varies harmonically in time:

$$F_{\text{ext}} = F_0 \cos \omega t,$$

with some frequency ω (the frequency of the natural free oscillations of the system will now be denoted by ω_0). Under the action of this force, oscillations occur in the system, at the frequency ω of the variation of the force; these are called *forced oscillations*. The motion of the system will then be, in general, a superposition of both oscillations: the natural oscillations with frequency ω_0 , and the forced oscillations with frequency ω .

The natural oscillations have already been discussed; let us now consider the forced oscillations, and determine their amplitude. We write these oscillations in the form

$$x = B \cos (\omega t - \beta),$$

where B is the amplitude and β some as yet unknown phase difference between the external force and the oscillations which it causes. We have written β with a negative sign, i.e. as a lag in phase, because it will be found below that this is what in fact occurs.

The acceleration w of a body executing forced oscillations is determined by the simultaneous action of three forces: the restoring force $-kx$, the external force F_{ext} , and the friction $F_{\text{fr}} = -bv$. Hence

$$mw = -kx - bv + F_{\text{ext}}$$

Dividing this equation by the mass m , using the relation $k/m = \omega_0^2$, and again putting $b/m = 2\gamma$, we have

$$w = -\omega_0^2 x - 2\gamma v + F_{\text{ext}}/m.$$

We shall now use a convenient graphical method of representing the oscillations, based on the fact that $x = B \cos \phi$ (where ϕ is the phase of the oscillation) may be geometrically regarded in an auxiliary *vector diagram* as the projection on a horizontal axis of a radius vector of length B at an angle ϕ to this axis. [To avoid misunderstanding it should be stressed that these “vectors” are not related to the concept of a vector as a physical quantity.]

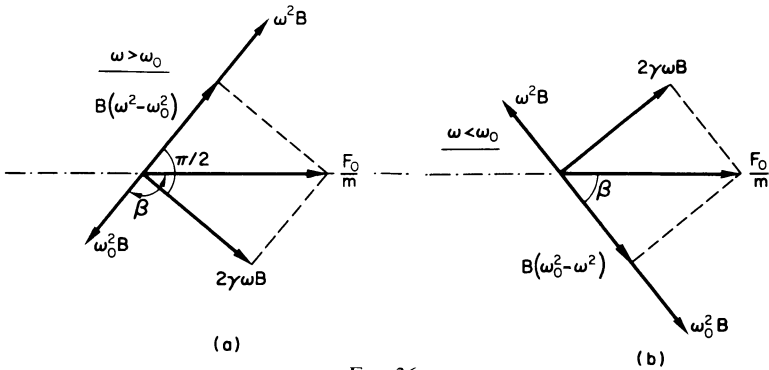


FIG. 36.

Each term in the above equation is a periodically varying quantity; the frequency ω is the same for each term but the phases are different. Let us consider, for example, the instant $t = 0$, when the phase of the external force $F_{\text{ext}} = F_0 \cos \omega t$ is zero, and so the quantity F_{ext}/m is represented by a horizontal vector of length F_0/m (Fig. 36). The quantity $\omega_0^2 x = \omega_0^2 B \cos(\omega t - \beta)$ oscillates with a phase lag of β ; it is represented by a vector of length

$\omega_0^2 B$ turned clockwise through an angle β relative to the force vector. The acceleration w has (as we have seen in §32) an amplitude $\omega^2 B$ and the opposite sign to the coordinate x ; it is therefore represented by a vector in the direction opposite to x . Finally, the velocity v has amplitude ωB and leads x by $\frac{1}{2}\pi$ in phase; the quantity $2\gamma v$ is represented by a vector of length $2\gamma\omega B$ perpendicular to the vector x .

According to the equation

$$F_{\text{ext}}/m = w + \omega_0^2 x + 2\gamma v,$$

the oscillation of F_{ext}/m must equal the sum of the oscillations of the three terms on the right-hand side. In the graph this means that the sum of the horizontal projections of these three vectors must equal F_0/m . For this to be so, the vector sum of these vectors must evidently be equal to the vector F_{ext}/m . The diagram (which gives separately the cases $\omega > \omega_0$ and $\omega < \omega_0$) shows that this is true if

$$(2\gamma\omega B)^2 + B^2(\omega^2 - \omega_0^2)^2 = (F_0/m)^2.$$

Hence we find the required amplitude of the oscillations:

$$B = \frac{F_0/m}{\sqrt{[(\omega^2 - \omega_0^2)^2 + 4\gamma^2\omega^2]}}.$$

The same diagrams can also be used to determine the phase difference β ; the expression for it will not be written out here, but it may be noted that the angle of lag of the oscillations of x relative to the external force is acute or obtuse according as $\omega < \omega_0$ or $\omega > \omega_0$.

We see that the amplitude of the forced oscillations is proportional to that of the external force F_0 , and also depends on the relation between the frequency ω of this force and the natural frequency ω_0 of the system. The amplitude of the oscillations reaches its maximum value when these two frequencies are equal ($\omega = \omega_0$); this is called *resonance*. The maximum value is

$$B_{\text{max}} = F_0/2m\omega_0\gamma,$$

and is inversely proportional to the damping coefficient γ . For this

reason the friction in the system cannot be neglected at resonance, even if it is slight.

It is interesting to compare the value B_{\max} with the static displacement B_{sta} that the body would have under a constant force F_0 . This displacement can be obtained from the general formula for B by putting $\omega = 0$: $B_{\text{sta}} = F_0/m\omega_0^2$. The ratio of the resonance displacement to this static displacement is

$$B_{\max}/B_{\text{sta}} = \omega_0/2\gamma.$$

We see that the relative increase in the amplitude of the oscillations in resonance (as compared with the static displacement) is given by the ratio of the frequency of natural oscillations to the damping coefficient. For systems with small damping this ratio may be very large. This explains the very great importance of the phenomenon of resonance in physics and technology. It is widely utilised in order to amplify oscillations, and always avoided if the resonance may cause an undesirable increase in the oscillations.

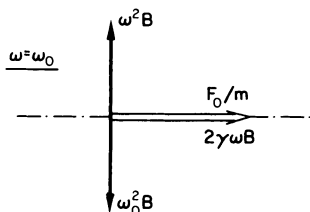


FIG. 37.

The origin of the amplification of the oscillations by resonance may be understood by considering the relation between the phases of the external force F_{ext} and the velocity v . When $\omega \neq \omega_0$ there is a difference of phase, and therefore the force F_{ext} is in the opposite direction to the velocity during a certain fraction of each period, and there is then a tendency for the motion to be retarded instead of accelerated. At resonance, however, the phases of the force and the velocity are the same (see the vector diagram, Fig. 37); thus the force always acts in the direction of the motion and continually “pushes” it.

Near resonance (i.e. when the difference $|\omega - \omega_0|$ is small in comparison with the resonance frequency ω_0), the formula for

the amplitude of the forced oscillations can be simplified. Putting in the denominator $\omega^2 - \omega_0^2 = (\omega + \omega_0)(\omega - \omega_0)$, we can approximately replace the sum $\omega + \omega_0$ by $2\omega_0$, and also replace ω by ω_0 in the term $4\gamma^2\omega^2$. The result is

$$B = \frac{F_0}{2m\omega_0\sqrt{[(\omega - \omega_0)^2 + \gamma^2]}}$$

This formula may also be written

$$B = \frac{\gamma B_{\max}}{\sqrt{[(\omega - \omega_0)^2 + \gamma^2]}}$$

where $B_{\max} = F_0/2m\omega_0\gamma$ is the maximum value of the amplitude at resonance.

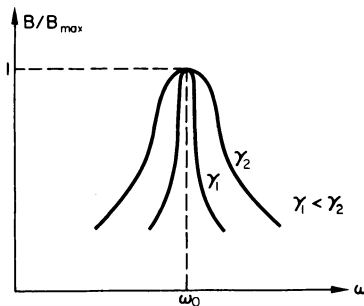


FIG. 38.

Figure 38 shows the resonance curves of amplitude as a function of frequency in accordance with this formula, for various values of the damping coefficient γ ; the ordinate is the ratio B/B_{\max} . So long as the absolute magnitude of the difference $\omega - \omega_0$ is small in comparison with γ , the amplitude B does not differ greatly from its maximum value; the amplitude begins to decrease considerably when $|\omega - \omega_0| \sim \gamma$. For this reason the "width" of the resonance curve is said to be of the order of γ . The height of the maximum (for given F_0) is inversely proportional to γ . Thus, the smaller the damping, the sharper is the resonance maximum, and the narrower is the resonance curve.

It has been mentioned above that the motion of an oscillating

system under the action of a periodic external force is a superposition of forced and natural oscillations. Neglecting the slight damping of the natural oscillations, we have a superposition of two simple harmonic oscillations, with frequencies ω and ω_0 and some amplitudes A and B . Near resonance, the frequencies ω and ω_0 are nearly equal, i.e. the difference $\Omega = |\omega - \omega_0|$ is small in comparison with ω and ω_0 . Let us find the nature of the resulting motion. To do so, we use a vector diagram in which each of the oscillations is represented by a vector (A and B in Fig. 39).

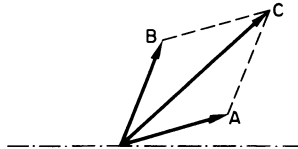


FIG. 39.

As the phases of the vibrations vary with time, these vectors rotate uniformly with angular velocities ω and ω_0 ; during one period T a vector makes one rotation, i.e. turns through an angle 2π , and its angular velocity is $2\pi/T$, which is just the angular frequency. The total oscillation is given by the geometric sum of the two vectors, the vector C . This vector, unlike A and B , has a length which varies with time, since the angle between the vectors A and B changes owing to the difference in the angular velocities ω and ω_0 . The length of C will evidently vary between $C_{\max} = A + B$, when the vectors A and B are in the same direction, and $C_{\min} = |A - B|$, when they are in opposite directions. This variation occurs periodically with frequency Ω , the latter being the relative angular velocity of rotation of the vectors A and B .

In the case considered, where the frequencies ω_0 and ω are almost equal, the vectors A and B rotate rapidly while at the same time having only a slow relative rotation. The variation of the resultant vector C may be regarded as a uniform rotation with the same frequency $\omega_0 \approx \omega$ (neglecting the difference between ω_0 and ω) together with a slow change in its length (with frequency Ω). In other words, the resultant motion is an oscillation with a slowly varying amplitude.

The periodic variation of the resultant amplitude in the superposition of oscillations with neighbouring frequencies is called

beats, and Ω is called the *beat frequency*. Figure 40 shows the beats when $A = B$.

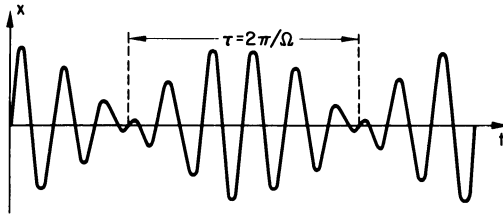


FIG. 40.

§36. Parametric resonance

Undamped oscillations can be caused not only by a periodic external force but also by a periodic variation of the parameters of the oscillating system. This is called *parametric resonance*. As an example, we may consider the build-up of the oscillations of a swing by a person who rhythmically stands up and sits down and thereby periodically changes the position of the centre of gravity of the system.

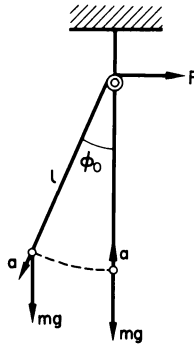


FIG. 41.

To elucidate the mechanism of this method of causing oscillations, let us take the simple example of a pendulum whose length can be varied by pulling and releasing a string on a pulley (Fig. 41). Let us suppose that, at each passage through the equilibrium (vertical) position, the pendulum is raised by the external force F through a short vertical distance a (small compared with the length l of the pendulum) and, at each extreme position, the string

is released the same distance a . During each period, therefore, the pendulum is twice lengthened and shortened, and so the frequency of the periodic variation of the parameter (the length of the pendulum) is twice that of the natural oscillations.

Since the string is lengthened when the pendulum is deflected, it will descend a distance $a \cos \phi_0$ (where ϕ_0 is the angular amplitude of the oscillations of the pendulum), which is less than the height a through which it rises when the string is raised. In each raising and lowering the external force acting on the string will do an amount of work

$$mga(1 - \cos \phi_0) \approx \frac{1}{2}mga\phi_0^2$$

against gravity (since ϕ_0 is assumed small, $\cos \phi_0 \approx 1 - \frac{1}{2}\phi_0^2$). In addition, the external force F does work against the centrifugal force which tightens the string. In the lowest position this work is mv_0^2/l (where v_0 is the maximum velocity of the pendulum) and in the extreme positions of the pendulum it is zero, since the velocity is zero. Thus the total work done by the external force in one period of oscillation of the pendulum is

$$A = 2(\frac{1}{2}mga\phi_0^2 + mv_0^2a/l).$$

But $v_0 = l\phi_0\omega$, where $\omega = \sqrt{g/l}$ is the frequency of the oscillations of the pendulum; thus

$$A = 6(a/l) \cdot \frac{1}{2}mv_0^2.$$

We see that the work done by the external force on the pendulum is positive and is proportional to the energy of the pendulum. This energy will therefore increase steadily, receiving in each period a small increment proportional to the energy and to the quantity a/l .

This is the mechanism of parametric resonance. A periodic variation of the parameters of an oscillating system (with a period twice the natural frequency of the system) may bring about a steady increase in its mean energy E , the rate of increase being proportional to E :

$$dE/dt = 2\kappa E,$$

where κ is a (small) constant. This is a relation of the same form as for damped oscillations, except that the derivative dE/dt is positive instead of negative. This means that the energy (and therefore the amplitude) of the oscillations increases exponentially with time.

In reality, of course, there is always some friction which causes damping of the oscillations. Consequently, in order for parametric resonance build-up of oscillations to occur, the amplification coefficient κ must exceed a certain minimum value, namely the damping coefficient due to friction.

We have discussed the production of oscillations in a system by periodic external interaction. There are, however, oscillatory systems in which oscillations are caused not by a periodic force but by a steady source of energy, which compensates the energy losses in the system that bring about the damping of the oscillations. One example of such a system is a clock, in which a compressed spring or raised weights act as the source of energy.

THE STRUCTURE OF MATTER

§37. Atoms

We shall not give here a detailed account of the problems of atomic physics, but simply describe some of the basic facts concerning the structure of matter which will be needed subsequently.

All bodies are made up of a fairly small number of simple substances, the *chemical elements*. The smallest particle of each element is an *atom* of that element.

The masses of the atoms are extremely small. It is therefore more convenient to measure them in special units, and not in grams. It would be natural to take as the unit the mass of the lightest atom, that of hydrogen. However, the precise standard of atomic weights is customarily taken not as the atom of hydrogen but as that of oxygen, which is more convenient for chemical purposes. The oxygen atom is approximately 16 times heavier than the hydrogen atom, and the unit of atomic weight is taken as one-sixteenth of the mass of the oxygen atom; this definition will be slightly refined in §38. The mass of the atom of any element expressed in these units is called the *atomic weight* of the element, and is usually denoted by A . The atomic weight of hydrogen is 1.008.

The mass of an atom in grams is proportional to its atomic weight. It is therefore clear that the number of atoms in a quantity of any element whose mass in grams is numerically equal to the atomic weight of that element (called a *gram-atom* of the element) is the same for every element. This is called *Avogadro's number*; its value is

$$N_0 = 6.02 \times 10^{23}.$$

The mass of an atom having atomic weight A is clearly

$$m_A = A/N_0 = 1.66 \times 10^{-24} A \text{ g.}$$

Although the atom is the smallest particle of an element, it has itself a complicated structure. An atom consists of a relatively heavy positively charged *nucleus* and a number of lighter negatively charged particles moving round it, the *electrons*, which form the *electron shell* of the atom. Different atoms have different nuclei, but all electrons are identical.

The mass of an electron is so much less than that of a nucleus that practically all the mass of an atom is concentrated in its nucleus. The lightest nucleus is that of hydrogen, called a *proton*, which is approximately 2000 times (more precisely, 1837 times) heavier than the electron. The absolute mass of the electron is

$$m = 9.11 \times 10^{-28} \text{ g.}$$

At the same time, the nucleus occupies only a negligible part of the volume of the atom. The radii of atoms, i.e. the radii of the regions round the nuclei in which the electrons move, are of the order of 10^{-8} cm; the radii of the nuclei are tens of thousands of times smaller, between 10^{-13} and 10^{-12} cm.

The charge on an electron is in absolute value

$$\begin{aligned} e &= 4.80 \times 10^{-10} \text{ CGSE unit of charge} \\ &= 1.60 \times 10^{-19} \text{ C.} \end{aligned}$$

It is often necessary to consider the electron charge multiplied by Avogadro's number, i.e. the charge on one "gram-electron". This product is called the *faraday*:

$$F = eN_0 = 9.65 \times 10^4 \text{ C.}$$

An atom as a whole is electrically neutral, its total charge being zero. In other words, the positive charge on the nucleus is exactly compensated by the negative charge on the surrounding electrons. This means that the charge on the nucleus is always an integral multiple of the charge on an electron. We can say that the magnitude of the charge on the electron is an *elementary charge*; the charge on any particle existing in Nature is a multiple of the electron charge. This is one of the most fundamental physical properties of matter.

The charge on a nucleus, expressed in units of the electron charge, is called the *atomic number* of the element and is usually

denoted by Z . Since the charge on the nucleus is just balanced by that on the electrons, the number Z is evidently also the number of electrons in the electron shell of the atom. All the properties of atoms that appear under ordinary conditions are determined by their electron shells, including, for example, the chemical and optical properties of matter. It is therefore clear that the atomic number is a fundamental characteristic of the atom, which to a considerable extent determines its properties. The arrangement of the elements in Mendeleev's periodic system is simply an arrangement in order of increasing atomic number, the latter being the same as the number giving the position of the element in the table.

The interaction forces which determine the structure of the atom are mainly those of electrical interaction of the electrons with the nucleus and with one another: the electrons are attracted by the nucleus and repelled by one another. Other forces (magnetic forces) are of relatively minor importance in the atom. The charge on the nucleus, and therefore the electric field in which the electrons move, are determined by the atomic number, thus again showing the fundamental importance of this number in governing the properties of atoms.

The gravitational interaction plays no part at all in atoms. The energy of the electrical interaction, for example, between two electrons at a distance r apart is e^2/r , and that of the gravitational interaction is Gm^2/r ; the ratio of these two quantities is

$$Gm^2/e^2 = 2.3 \times 10^{-43}.$$

This number is so small that it is pointless to speak of gravitational interaction in the atom.

The properties of atoms can in no way be described by means of classical mechanics, which is unable to explain either the structure of the atom or even the fact that it exists as a stable configuration. Classical mechanics is entirely inapplicable to the motions of particles of such small mass as electrons in such small regions of space as are occupied by atoms. Atomic phenomena can be understood only from the laws of a quite different mechanics called quantum mechanics.

Under various external interactions, an atom may lose one or more electrons from its electron shell. We then have an atomic

particle which is not electrically neutral but charged, a positively charged *ion*. The energy needed to remove one outer electron from the atom is called the *ionisation potential* of the atom.

Energies in atomic phenomena are usually measured in a special unit, since the erg would be too large for this purpose. The unit employed is the energy gained by an electron in traversing a potential difference of one volt in an electric field, and is called an *electron-volt* (eV). Since the work done by the electric field is equal to the product of the charge and the potential difference, and 1 volt is 1/300 CGSE unit of potential, we have

$$\begin{aligned} 1 \text{ eV} &= 4.80 \times 10^{-10} \times 1/300 \text{ erg} \\ &= 1.60 \times 10^{-12} \text{ erg.} \end{aligned}$$

The ionisation potentials of atoms are measured in electron-volts. They range from 3.89 eV, the smallest (for caesium), to 24.6 eV, the largest (for helium). The ionisation potential of the hydrogen atom is 13.6 eV.

If we consider the ionisation potential as a function of atomic number, we see that this function has a remarkable periodic form. It increases more or less steadily in each period of Mendeleev's table, reaching its greatest value for a noble-gas atom, and then drops sharply at the beginning of the next period. This is one of the principal manifestations of the periodic properties of atoms, which gave the periodic system its name.

The ionisation potential represents the binding energy of the outer electrons in the atom. The inner electrons, moving deep in the shell, have considerably higher binding energies. The energy which would be needed to remove the deepest electrons from the electron shell reaches 10^4 or 10^5 eV in heavy atoms.

As well as positively charged ions, negative atomic ions can also exist, in which an extra electron is attached to the atom. However, by no means all isolated atoms are capable of attaching an electron to form a stable system, i.e. not all atoms have an *affinity* for an extra electron. Only the halogen atoms (F, Cl, Br, I), hydrogen and the atoms of elements of the oxygen group (O, S, Se, Te) can form negative ions. These elements have different affinities for an electron; it is greatest for the halogens and least for hydrogen, where the binding energy in the negative ion is only about 0.1 eV.

Ions are usually denoted by the symbol of the chemical element with indices + or - equal in number to the charge on the ion: H^+ , Cl^- , etc. Sometimes dots and primes are used: H^\cdot , Cl' .

§38. Isotopes

The nuclei of atoms are composite structures, in general consisting of many particles. Their constituent parts are protons (hydrogen nuclei) and *neutrons*, which are particles whose mass is almost equal to that of the proton but which differ from protons in having no electric charge. The total number of protons and neutrons in the nucleus is called its *mass number*. Since the charge on the nucleus is determined by the protons present, its value in units of the elementary charge e is equal to the number of protons; that is, the number of protons in the nucleus is equal to the atomic number Z . The remaining particles in the nucleus are neutrons.

The particles in the nucleus are held together by specific forces, which are not electrical. These forces are extremely strong, and the binding energies of the particles in the nucleus are tens of millions of electron-volts, and so are very large in comparison with the binding energies of the electrons in the atom. For this reason the atomic nuclei undergo no internal changes in any phenomena which are not of nuclear origin, and behave simply as particles of given mass and charge.

We have already mentioned that the properties of an atom are determined mainly by the charge on the nucleus. The mass of the nucleus is of relatively minor importance. This fact is clearly evident in atoms with the same atomic number but having nuclei of different masses. It is found that the atoms of a given chemical element are not all identical; though they have the same number of electrons, their nuclei may differ in mass while having the same charge. Such varieties of a given element are called *isotopes* of the element. All the isotopes of an element are chemically identical, and their physical properties are also very similar. The number of isotopes existing in Nature is different for different elements, varying from one (as in Be, F, Na, Al, etc.) to ten in tin.†

†To avoid misunderstanding it should be emphasised that this refers only to the isotopes found in Nature. Other isotopes may be produced artificially, but their nuclei are unstable and disintegrate spontaneously.

The elements found on the Earth are mixtures of various isotopes in definite proportions. The atomic weights given in a table of chemical elements are the mean weights of the atoms in these mixtures (often referred to as chemical atomic weights) and not the exact weights of any particular isotopes. The atomic weights of the isotopes are very close to whole numbers, the mass numbers, differing from them only by a few parts in a hundred or even a few parts in a thousand. The mean (chemical) atomic weights, on the other hand, of course need not be whole numbers.

It is therefore necessary to refine the definition of the unit of atomic weight as one-sixteenth of the atomic weight of oxygen, given in §37. Oxygen has three isotopes: ^{16}O , ^{17}O and ^{18}O . (The atomic weight, or more precisely the mass number, of an isotope is customarily written as an index to the left of the chemical symbol of the element.) The most abundant of these isotopes is ^{16}O ; the isotopes ^{17}O and ^{18}O occur in the natural mixture only in amounts of 0.04 and 0.2% respectively. Although these are relatively small quantities, they are important in a precise definition of atomic weights.

The mean atomic weights of the natural mixtures of isotopes are usually defined relative to the atomic weight of natural oxygen, taken as exactly 16; this is sometimes called the chemical scale of atomic weights. In order to define the precise atomic weights of individual isotopes in nuclear physics, it is natural to use the atomic weight of one particular isotope as basis; the atomic weight of the isotope ^{16}O is taken as 16, and the unit of this scale (the physical scale) is 0.027% less than that of the chemical scale.

The use of two scales of atomic weights involves some inconvenience, and it has therefore been recommended that a single new scale should be employed, in which atomic weights are defined relative to that of the carbon isotope ^{12}C , taken as 12. This change means only a very slight increase (by 0.0043%) in the ordinary chemical atomic weights.

The first element in the periodic system, hydrogen, has two natural isotopes. As well as the principal isotope of atomic weight 1, there is an isotope of weight 2; natural hydrogen contains only one ^2H atom to about 6000 ^1H atoms. The "heavy" isotope of hydrogen is usually denoted by a separate symbol D

and called *deuterium*; its nucleus is called a *deuteron*. Since the ratio of masses of the two isotopes of hydrogen is 2, which is comparatively large, the difference between their physical properties is greater than for isotopes of other elements, where the relative mass difference is considerably less. For example, "heavy water" D_2O , containing the heavy isotope deuterium, freezes at $3.8^\circ C$ instead of $0^\circ C$, and boils at $101.4^\circ C$ instead of $100^\circ C$.

It may also be mentioned that the next element, helium, likewise has two isotopes, 3He and 4He . The isotope 4He is by far the more abundant, atoms of 3He being present in natural helium only to the extent of 1 to about 10^6 atoms of 4He . The isotope 3He can, however, be artificially prepared in large quantities by the methods of nuclear physics.

§39. Molecules

Atoms of different elements can combine with one another to form *molecules*. The forces of interaction between atoms which bring about the formation of molecules (called chemical interaction), like the forces acting within the atom itself, are fundamentally electrical. But the formation of molecules, like the structure of atoms, is a quantum phenomenon, and cannot be explained in terms of classical mechanics. Here we shall describe only the basic properties of this interaction, without investigating its nature.

The simplest molecule is a diatomic molecule, consisting of two like or unlike atoms. The interaction of atoms which causes the formation of such a molecule is described by a potential energy

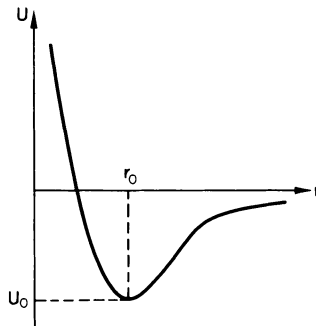


FIG. 42.

having the form shown diagrammatically in Fig. 42. In this diagram the potential energy U of the interaction between the two atoms is plotted as a function of the distance r between them (more precisely, of the distance between their nuclei). This function has a fairly deep and sharp minimum at a certain point $r = r_0$. At smaller distances the curve rises very steeply; this region corresponds to a strong repulsion between the atoms, due essentially to the Coulomb repulsion between the nuclei as they approach. At large distances the atoms attract each other.

The distance r_0 corresponds to a stable equilibrium position of the nuclei in the molecule. In reality, the nuclei do not occupy exactly these positions, but oscillate about them; the amplitude of these oscillations, however, is usually small. The depth U_0 of the potential well represents the strength of the bond between the atoms in the molecule; strictly speaking, the precise value of the *binding energy* needed to separate the atoms is slightly different from U_0 because of the energy of the oscillations of the nuclei.

The table below gives as an illustration the values of r_0 in ångströms ($1 \text{ \AA} = 10^{-8} \text{ cm}$) and U_0 in electron-volts for some diatomic molecules.

Molecule	r_0	U_0
H ₂	0.75	4.5
O ₂	1.2	5.1
Cl ₂	2.0	2.5
N ₂	1.1	7.4

A diatomic molecule may be likened to a dumb-bell of length r_0 . Polyatomic molecules have a more complex form. Figure 43 shows the positions of the atomic nuclei in some triatomic molecules, the distances being shown in ångströms. Some of these form triangles (water H₂O and ozone O₃), while in others the atoms are collinear (carbon dioxide CO₂ and hydrocyanic acid HCN). In §40 examples of still more complicated molecules will be given.

We see that the distances between the nuclei in molecules are of the order of 10^{-8} cm , like the dimensions of the atoms themselves. Thus the atoms in a molecule are close together, and it

is therefore, strictly speaking, impossible to distinguish the electron shells of the individual atoms in a molecule. Although the internal regions of the electron shells are not much changed when atoms combine to form a molecule, the motion of the outer electrons may be considerably modified, and these electrons are as it were shared by the atoms.

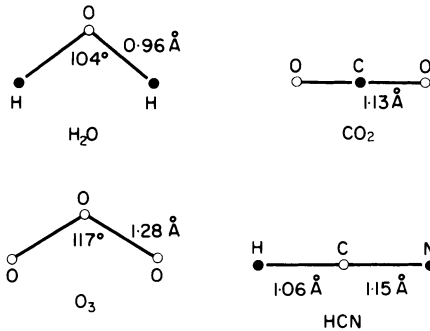


FIG. 43.

In some molecules the outer parts of the electron shells are rearranged so that there are fewer electrons on average round some nuclei and more round others than in the neutral atoms; such molecules may be regarded as consisting of ions (for example, the KCl molecule may be regarded as consisting of the positive ion K^+ and the negative ion Cl^-). In other cases (e.g. H₂, O₂, HCl) the atoms in the molecule remain on average electrically neutral. This difference, however, is only quantitative, and various intermediate cases between the two limiting ones mentioned are possible.

A characteristic property of the chemical interaction of atoms is that it can be *saturated*. This means that the atoms which combine with one another because of this interaction cease to be able to interact in the same way with other atoms.

Different molecules also interact; this interaction is called the *van der Waals interaction*, to distinguish it from the chemical interaction of atoms which leads to the formation of molecules.

The interaction of molecules can not in general be represented simply by a curve $U = U(r)$ as was done above for atoms, since the relative position of the molecules is described by a larger

number of parameters: as well as the distance r between the molecules, their relative orientation is important. But if the interaction between molecules is regarded as being averaged over all possible orientations, it can again be represented by such a curve. This curve resembles the curve of interaction between atoms in a molecule, in that at large distances all molecules attract one another, and at small distances repel one another. The forces of attraction between molecules decrease rapidly with increasing distance between them. The forces of repulsion between molecules increase even more rapidly as they approach, so that approaching molecules behave like solid bodies and do not interpenetrate. The depth of the minimum on the van der Waals interaction curve is very small, being only some tenths or even hundredths of an electron-volt (see §68), whereas the depth of the potential well on the curve of chemical interaction is several electron-volts.

Another important difference between the two kinds of interaction is that the van der Waals forces, unlike the chemical forces, do not exhibit saturation. The van der Waals interaction exists between all molecules, so that if two molecules are brought together by it they continue to attract other molecules. The forces of molecular attraction therefore do not lead to the formation of "supermolecules", but merely assist the general tendency of all molecules to approach one another. This tendency exists when matter enters a condensed (liquid or solid) state.

THE THEORY OF SYMMETRY

§40. Symmetry of molecules

The concept of symmetry plays a fundamental part in physics. Symmetry is one of the most important characteristics of a given physical object, and in many cases it has a decisive effect on the behaviour of the object.

We shall begin by considering the possible symmetries of individual molecules. The symmetry properties consist of various *symmetry elements*, which we shall first of all define.

A molecule is said to have an *axis of symmetry of order n* if it is left unchanged in position by a rotation through an angle $2\pi/n$ about that axis, where n is any integer: $n = 2, 3, 4, \dots$; such an axis is usually denoted by the symbol C_n . For example, if a molecule has an axis of symmetry of order 2, this means that the molecule is unchanged in position by a rotation through 180° ; that is, each atom A, B, \dots in the molecule corresponds to another atom A', B', \dots of the same kind, located as shown in Fig. 44 relative to A, B, \dots and the axis. If the molecule has an axis of symmetry of order 3, it is left unchanged in position by rotations through 120° and 240° ; for each atom A , the molecule also contains two atoms of the same kind, A' and A'' , situated as shown in Fig. 45.

A molecule has a *plane of symmetry* if it is left unchanged in position on reflection in that plane; this symmetry element is denoted by σ . This means that for each atom A in the molecule there is another atom A' of the same kind lying on the perpendicular from A to the plane and at the same distance on the other side of the plane (Fig. 46).

In addition to reflection in a plane, we can define "reflection in a point", giving a new symmetry element, a *centre of symmetry* or *centre of inversion*, denoted by i . If a molecule has a centre of symmetry at some point i (Fig. 47), each atom A corresponds to

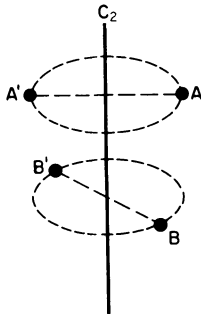


FIG. 44.

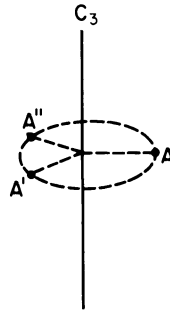


FIG. 45.

another atom A' of the same kind, lying on the line joining A and i and at the same distance on the other side of i .

Finally, there is a symmetry element called a *rotary-reflection axis of order n* , denoted by S_n . A molecule possesses this symmetry if it is unchanged in position by rotation through an angle $2\pi/n$ about an axis followed by reflection in a plane perpendicular to that axis. The order n of a rotary-reflection axis can only be even; if n is odd, say $n = 3$, then by repeating the rotary reflection 6 times we easily see that the axis S_n is in fact a combination of two independent symmetry elements, an axis of symmetry C_3 and a plane of symmetry σ perpendicular to it. If a molecule has a rotary-reflection axis of order 4, for example, then each atom A corresponds to three other atoms A' , A'' , A''' of the same kind, arranged as shown in Fig. 48. The presence of such an axis necessarily involves the presence of a simple axis of symmetry, of half the order (in this case, C_2).

[It may be noted that, as is easily seen, a rotary-reflection axis of order 2 is equivalent to a centre of symmetry at the point where the axis meets the plane of reflection. Thus S_2 is not a new symmetry element.]

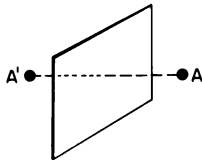


FIG. 46.

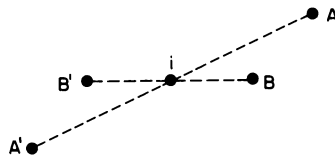


FIG. 47.

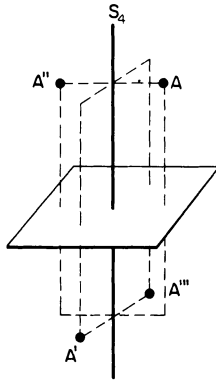


FIG. 48.

These are the elements of which the symmetry of a molecule may be composed. The following are examples of the way in which various combinations of these elements occur to determine the symmetry of a molecule.

The water molecule H_2O forms an isosceles triangle (Fig. 49). Its symmetry consists of an axis of order 2 (the altitude of the triangle) and two mutually perpendicular planes of symmetry passing through this axis C_2 .

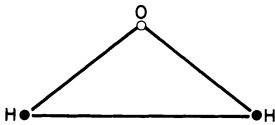


FIG. 49.

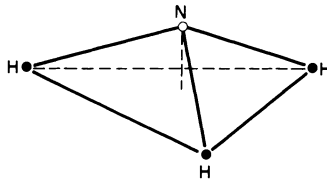


FIG. 50.

The ammonia molecule NH_3 forms an equilateral triangular pyramid with the N atom at the vertex and the H atoms at the corners of the base (the pyramid is actually very flat, its altitude being only about $\frac{1}{4}$ of the edge of the base). The symmetry consists of a vertical axis of order 3 (Fig. 50) and three planes of symmetry passing through this axis at angles of 60° , each plane passing through one of the H atoms.

A still greater number of symmetry elements occurs in the benzene molecule C_6H_6 , whose atoms lie in a plane and form a

regular hexagon (Fig. 51). The plane containing the atoms is obviously a plane of symmetry, and the molecule also has an axis of symmetry of order 6 passing through the centre of the hexagon at right angles to the plane. The centre of the hexagon is a centre of symmetry. There are also six axes of order 2, three of which join diametrically opposite atoms while the other three bisect opposite sides of the hexagon; one axis of each kind is shown in Fig. 51. Finally, the six planes through these axes C_2 at right angles to the plane of the diagram are a further six planes of symmetry.

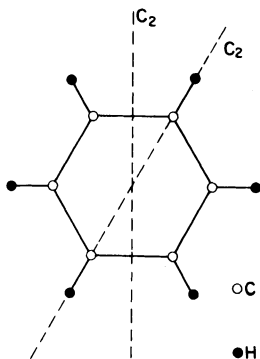


FIG. 51.

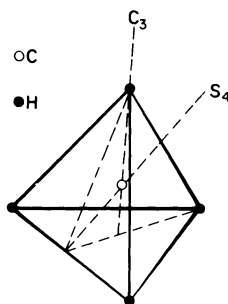


FIG. 52.

Let us consider also the methane molecule CH_4 , which is a regular tetrahedron (a solid with four equal faces, each an equilateral triangle): the H atoms are at the four vertices, and the C atom at the centre (Fig. 52). This molecule has four axes of symmetry of order 3, each passing through one vertex and the centre of the tetrahedron. Three rotary-reflection axes of symmetry of order 4 pass through the midpoints of opposite edges of the tetrahedron. Finally, there are six planes of symmetry, each passing through one edge and the midpoint of the opposite edge. Figure 52 shows one of each of these symmetry elements.

§41. Stereoisomerism

There is a curious effect which depends on the presence or absence of a certain degree of symmetry in the molecule. If a sufficiently asymmetric molecule undergoes a mirror reflection, the resulting molecule is similar to the first, but not identical

with it. The molecule CHClBrI is of this type, for example; it is obtained by replacing three H atoms in the methane molecule CH_4 by three different atoms Cl, Br and I. Figure 53 shows two such molecules which are mirror images in a vertical plane. It is obvious that these two molecules cannot be made to coincide by any rotation in space, and in this sense they are not identical.

Two such similar but not identical molecules derived from each other by reflection are called *stereoisomers*. One isomer is said to be right-handed and the other left-handed.

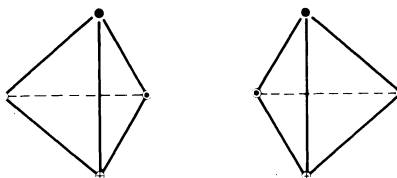


FIG. 53.

By no means all molecules can have stereoisomers. The existence of these depends on the symmetry of the molecule. For example, if a molecule has even one plane of symmetry, its mirror image is identical with it, and they differ only by a rotation about some axis. Hence, for example, there are no stereoisomers, not only of the highly symmetric molecule CH_4 , but even of the much less symmetric molecules CH_3Cl and CH_2ClBr , which still possess a plane of symmetry.

Similarly, molecules which have a centre of symmetry, or any rotary-reflection axis, have no stereoisomers.

Stereoisomers are completely identical in almost all physical properties. The difference between them appears, in particular, in certain phenomena which occur when light passes through solutions of these substances (for which reason stereoisomers are also called *optical isomers*).

The difference between stereoisomers has an important effect when they react with other molecules which are also asymmetric. The reaction between right-handed isomers of the two substances occurs in the same way as that between their left-handed isomers: the two processes differ only by a mirror reflection and so must be identical in physical properties. Similarly, the reaction of a right-handed isomer with a left-handed one is the same as

that of a left-handed isomer with a right-handed one. But the course of the reaction in the two latter cases differs considerably from that in the two former cases. This is the principal difference between stereoisomers.

If asymmetric molecules are formed in a chemical reaction between two symmetric substances (not having stereoisomers), then, since the initial substances are unchanged by reflection, the same must be true of the product. The reaction therefore yields a mixture of equal quantities of the two isomers.

§42. Crystal lattices

The fundamental property of crystals is that their atoms are regularly arranged. It is this symmetry of the internal arrangement of the atoms in crystals which we shall discuss, and not that of the external shape of crystals.

The set of points at which the atoms (or more precisely the atomic nuclei) are located is called a *crystal lattice*. In considering the symmetry of the lattice we may regard it as unbounded in space, ignoring the boundaries of the crystal, since these do not affect the structure of the lattice as such.

The fundamental characteristic of a crystal lattice is the spatial periodicity of its structure: the crystal as it were consists of repeated units. The lattice may be divided by three families of parallel planes into identical parallelepipeds containing equal numbers of atoms arranged in the same manner. The crystal is an assembly of such parallelepipeds in parallel positions. Thus, if the lattice is moved as a whole, parallel to itself, along the direction of any edge of the parallelepipeds through a distance equal to an integral number of times the length of that edge, the lattice will be unchanged in appearance. Such a displacement is called a *translation*, and the symmetry of the lattice with respect to these displacements is called *translational symmetry*.

The smallest parallelepiped which can be repeated to form the entire crystal lattice is called the *unit cell* of the lattice. The size and shape of this cell and the arrangement of the atoms in it completely determine the structure of the crystal. The lengths and directions of the three edges of the unit cell define three vectors called the *basic vectors* of the lattice; these are the shortest distances through which the lattice can be displaced so as to remain unchanged in appearance.

If there is an atom at a vertex of a unit cell, there must evidently be atoms of the same kind at every vertex of every cell. The assembly of these like atoms in corresponding positions is called a *Bravais lattice* of the crystal in question (Fig. 54). It is a kind of skeleton of the crystal lattice, displaying the whole of the translational symmetry, i.e. the complete periodicity. Any of its atoms can be moved to the position of any other by some translation of the lattice.

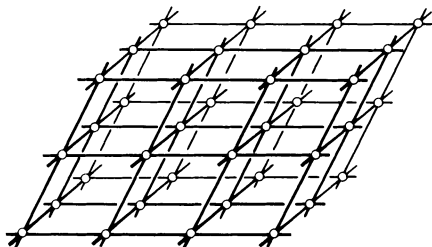


FIG. 54.

It should not be supposed that the atoms in a Bravais lattice are necessarily all the atoms in the crystal. They need not even be all the atoms of one kind. This important fact may be illustrated by an example, considering for clarity not a three-dimensional lattice (as the crystal lattice really is) but a two-dimensional one which is more easily shown in a diagram.

Let the lattice consist of only one kind of atom, represented by the dots in Fig. 55a. It is easy to see that, although all these atoms are of the same kind, they are not crystallographically equivalent. For the fact that all the atoms in a Bravais lattice are in corresponding positions means that, if any atom in it has a neighbour at a certain distance in a given direction, then all the atoms in the Bravais lattice will have neighbours of the same kind at the same distance in the same direction. It is therefore clear that the points of type 1 in Fig. 55a are not in positions of the same type as the points of type 2. The point 1 has a neighbour 2 at a distance d , but the point 2 does not have a neighbour at that distance in the same direction. The points 1 and 2 are therefore not equivalent and do not belong to the same Bravais lattice. Points of each type separately form a Bravais lattice at a distance d from that formed by points of the other type.

If the atoms 2 are moved to the centres of the squares formed by the points 1 (Fig. 55b), then all the atoms become equivalent; an atom 2 will have a neighbour 1 at the same distance and in the same direction as an atom 1 has a neighbour 2. In this structure all the atoms together form a single Bravais lattice.

It is clear from the foregoing that a crystal in general consists of several Bravais lattices which interpenetrate one another. Each of these corresponds to a particular type and arrangement of atoms, and all the lattices, regarded simply as sets of points, are identical.

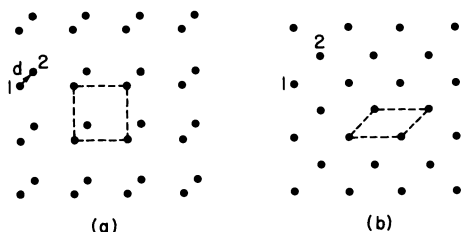
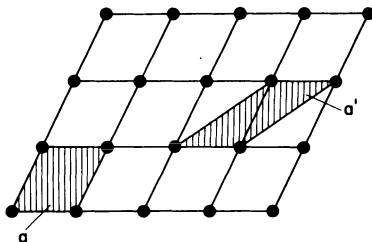


FIG. 55.

If all the atoms in a crystal form a single Bravais lattice, then each unit cell contains only one atom. For example, in Fig. 55b each cell (a parallelogram in the two-dimensional lattice) contains a single atom 1 or 2. [Here it may be noted that in counting the number of atoms per cell, only one vertex of each cell is to be taken, the remainder being assigned to adjoining cells.]

If, however, the crystal lattice is composed of several Bravais lattices, the unit cell contains more than one atom, namely one from each Bravais lattice. For example, in the lattice shown in Fig. 55a the unit cell contains two atoms: one atom 1 and one atom 2.



The division of the crystal into basic parallelepipeds, i.e. unit cells, is not unique. A unit cell may, in principle, be chosen in an infinity of ways. To illustrate this, let us again consider a two-dimensional lattice (Fig. 56). We can clearly regard the unit cell either as the parallelogram a or as the parallelogram a' with equal validity.

It is important, however, that despite this ambiguity the unit cell, however chosen, will contain the same number of atoms and have the same volume (in a two-dimensional lattice, the same area: the parallelograms a and a' have the same base and the same height, and therefore the same area). For let us consider atoms of a given kind and position. It is clear from the foregoing that each cell contains one such atom, and the number of cells in a volume V of the crystal is therefore always equal to the number N of these atoms. Thus the volume of one cell $v = V/N$, however the cell is chosen.

§43. Crystal systems

The Bravais lattice is a very important characteristic of a crystal, and the classification of the various types of crystal symmetry is based in the first instance on the classification of the various types of Bravais lattice.

All Bravais lattices have translational symmetry. In addition they may also have the symmetry elements discussed in §40, i.e. various axes and planes of symmetry. This symmetry is the basis of the classification described below.

For example, every point of a Bravais lattice is a centre of symmetry, since to each atom in the lattice there corresponds another atom collinear with that atom and the lattice point considered and at the same distance from this point. Thus any Bravais lattice has a centre of symmetry, but it may also possess higher symmetry.

A body of finite size, such as a molecule, may in principle have an axis of symmetry of any order. A periodic structure, on the other hand, such as a crystal lattice, can have axes of symmetry only of a small number of orders: 2, 3, 4 or 6. For if the lattice had an axis of symmetry of order 5, say, this would mean that the lattice contained a plane in which there were points forming regular pentagons. But this is certainly impossible; the only regular polygons which can completely fill a plane are equilateral

triangles, squares, and regular hexagons. In order to prove this statement, let us consider any point in the plane at which the sides of polygons filling the plane meet. In order to fill the plane completely, the angle between adjoining sides of a polygon must be an integral submultiple of 2π , i.e. must be $2\pi/p$, where p is any integer. The angle in a regular n -gon is $\pi(n-2)/n$. Thus we have

$$\frac{\pi(n-2)}{n} = \frac{2\pi}{p},$$

whence it is seen that the quantity $2n/(n-2)$ must be an integer, and this is true only for $n = 3, 4$, and 6 .

Thus we see that by no means all types of symmetry are possible in lattices. In consequence, there are only a relatively small number of types of symmetry of Bravais lattices. These are called *crystal systems*, and they will now be enumerated.

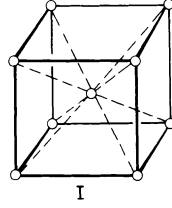
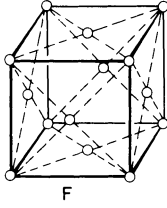
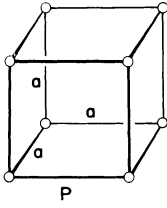
1. *Cubic system.* The most symmetrical Bravais lattice is one having the symmetry of a cube. (Instead of listing the axes and planes of symmetry of the lattice, we shall simply state the geometrical figure, in this case a cube, which has the same symmetry.)

This lattice is obtained by placing atoms at the vertices of cubic cells. There are, however, other ways of constructing a Bravais lattice with the symmetry of a cube. It is evident that the cubic symmetry is unaffected by placing an atom also at the centre of each cubic cell; all the atoms (at the vertices and at the centres of the cubic cells) will have the same relative position, i.e. the same neighbours, and will therefore form a single Bravais lattice. We can also construct a cubic Bravais lattice by adding to the atoms at the vertices of the cubic cells an atom at the centre of every face of the cubes.

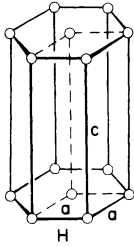
Thus there are three different Bravais lattices belonging to the cubic system. They are called *simple*, *body-centred* and *face-centred* lattices and denoted by the symbols P , I , F respectively. Figure 57 shows the arrangement of the atoms in the cells of these lattices.

The cubic cell of the simple Bravais lattice is also the unit cell, but the cubic cells of the lattices I and F are not unit cells, as we see from the fact that they contain more than one atom. In

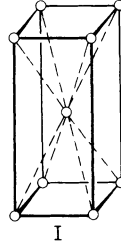
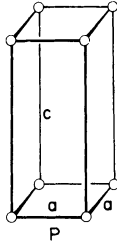
Cubic system



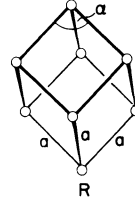
Hexagonal system



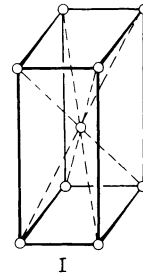
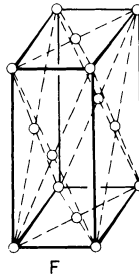
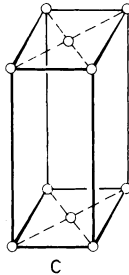
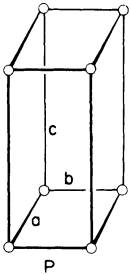
Tetragonal system



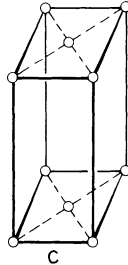
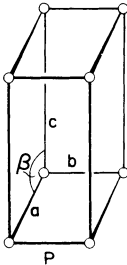
Rhombohedral system



Orthorhombic system



Monoclinic system



Triclinic system

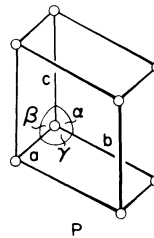


FIG. 57.

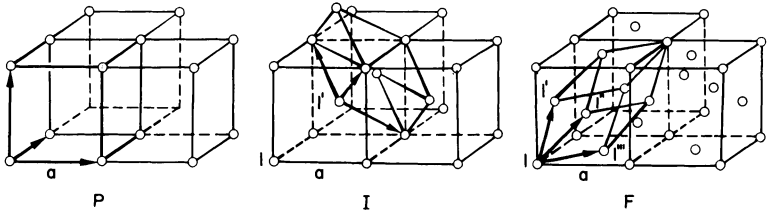


FIG. 58.

Fig. 58 the thick lines show the unit cells of all three types of cubic lattice. In the body-centred cubic cell there are two atoms (e.g. 1 and 1' in Fig. 58), and in the face-centred cell there are four atoms (1, 1', 1'', 1''' in the diagram); the other atoms must be regarded as belonging to adjoining cells. Hence it follows that the volumes of the unit cells in the body-centred and face-centred lattices are respectively $\frac{1}{2}a^3$ and $\frac{1}{4}a^3$, where a is the length of the edge of the original cube.

The length a is called the *lattice constant*. It is the only numerical parameter that is needed to describe a cubic lattice.

The unit cells in the body-centred and face-centred lattices have a form which does not itself possess the cubic symmetry of the lattice. In this sense the representation of the structure of the crystal by means of such cells does not exhibit its symmetry so clearly as the representation by means of the cubic cells which are not unit cells. The arrangement of atoms in the crystal is therefore usually described in terms of the cubic cells, using rectangular coordinates with axes X , Y , Z along three edges of the cubic cell and the constant a as the unit of measurement of the coordinates. For example, an atom at the centre of a cubic cell is described by the three coordinates $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$; the coordinates $\frac{1}{2}, \frac{1}{2}, 0$ define an atom at the centre of a face in the plane XY , and so on.

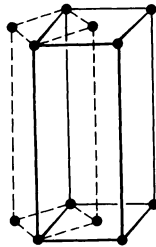


FIG. 59.

2. *Tetragonal system.* If a cube is stretched in the direction of one of its edges, a less symmetrical geometrical figure is obtained, namely a right square prism, whose symmetry corresponds to that of Bravais lattices of the tetragonal system.

There are two types of such lattices, simple and body-centred, whose cells are shown in Fig. 57. At first sight it appears that we could construct a lattice with the same symmetry by adding to the simple lattice cell one atom at the centre of each end face of the prism (Fig. 59). However, it is easily seen that such a lattice would be reduced to another simple tetragonal Bravais lattice by simply taking a different basic square prism cell, so that no new lattice results. For by joining the atoms at the centres of the end faces of two adjoining cells to the atoms at the vertices, as shown in Fig. 59, we obtain another prism whose symmetry is the same as the original one and which has atoms only at the vertices. Similarly, there is no face-centred tetragonal Bravais lattice, since it is equivalent to the body-centred one.

The tetragonal lattice is described by two constants, the edge length a of the base and the height c of the prism.

3. *Orthorhombic system.* If a cube is stretched along two of its edges by different amounts, we obtain a rectangular parallelepiped with edges of three different lengths. The symmetry of this figure is that of lattices of the orthorhombic system.

There are four types of orthorhombic Bravais lattices: simple, body-centred, face-centred and base-centred, the last of these being denoted by the symbol C . Fig. 57 shows, as for the other systems, the basic parallelepipeds of the orthorhombic lattices, whose forms correspond to the full symmetry of this system; here again, they coincide with the unit cell only in the simple Bravais lattice.

The orthorhombic lattice is described by three parameters, the lengths a , b , c of the edges of the cell. These are taken as the units of length on the axes of rectangular coordinates along the corresponding edges of the cell.

4. *Monoclinic system.* This has an even lower symmetry, namely that of the figure obtained from a rectangular parallelepiped by "slanting" it along one edge, giving a right parallelepiped with arbitrary base. This system includes two Bravais lattices, P and C in Fig. 57.

The monoclinic lattice is described by four parameters, the

lengths a, b, c of the edges of the cell and the angle β between two of these edges (the others being at right angles). Here again, the positions of the atoms are specified by means of coordinates with axes along the three edges of the cell, but these coordinates are now oblique and not rectangular.

5. *Triclinic system.* This corresponds to the symmetry of an arbitrary oblique parallelepiped. It is the lowest symmetry, comprising only a centre of symmetry. It includes only one Bravais lattice P , described by the lengths a, b, c of three edges of the cell and the angles α, β, γ between them.

Two further crystal systems stand somewhat apart.

6. *Hexagonal system.* The lattices of this system have very high symmetry, corresponding to that of a right regular hexagonal prism. The Bravais lattice of this system (denoted by H) can be constructed in only one way: its lattice points are at the vertices of hexagonal prisms and at the centres of their hexagonal faces.

The hexagonal lattice is described by two parameters: the edge length a of the base and the height c of the prism. The unit cell in this lattice is a parallelepiped whose base is a rhombus, as shown by the broken lines in Fig. 57. The edges of this unit cell (height c and two sides a of the base at an angle of 120°) are used as coordinate axes in specifying the position of the atoms in the lattice.

7. *Rhombohedral system.* This corresponds to the symmetry of a rhombohedron, a figure obtained by stretching or compressing a cube along one of its spatial diagonals without changing the length of the edges. All its faces are equal rhombuses. In the only Bravais lattice possible in this system (denoted by R), the lattice points are at the vertices of rhombohedra. This lattice is described by two parameters: the length a of the edges of the cell and the angle α between them. For $\alpha = 90^\circ$ the rhombohedron becomes a cube.

This completes the list of the various Bravais lattices. We see that there are altogether seven types of symmetry of the Bravais lattice, i.e. seven crystal systems, corresponding to fourteen different types of Bravais lattice.

The crystal systems are the basis of the classification of crystals and are principally used to describe the properties of the crystal. The terms "hexagonal crystal", "cubic crystal" and so on frequently used for brevity must be taken as indicating the crystal

system and not, for example, the external form of a particular specimen.

It may also be mentioned that crystals of the rhombohedral, hexagonal and tetragonal systems, whose lattices are described by two parameters, are called *uniaxial* crystals, while those of the triclinic, monoclinic and orthorhombic systems are called *biaxial* crystals.

§44. Space groups

The Bravais lattices discussed above are sets of atoms which are equivalent, i.e. of the same kind and similarly situated. It has already been stressed that a Bravais lattice does not in general include all the atoms in a crystal, and an actual crystal lattice can be represented as an assembly of several interpenetrating Bravais lattices. Although all these lattices are entirely identical, the symmetry of the assembly, i.e. the symmetry of the crystal itself, may differ considerably from that of a single Bravais lattice.

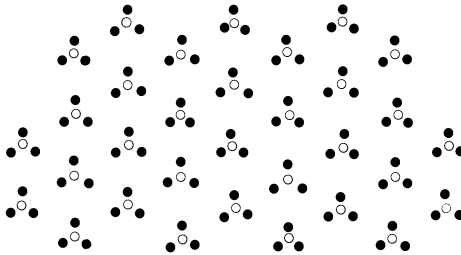


FIG. 60.

This important fact may be illustrated by an example, again using for clarity a representation of a two-dimensional lattice. In Fig. 60 the white circles are the points of a two-dimensional “hexagonal” Bravais lattice. An axis of symmetry of order 6 passes through each point of this lattice at right angles to the plane of the diagram. Now let three further lattices of the same kind be superposed on this lattice; their points are shown by the black circles in Fig. 60. It is clear that in the resulting lattice the axes of symmetry just mentioned will be of order 3 and not 6.

We see that, when the actual lattice is composite, its symmetry may be lower than that of its Bravais lattice.

In actual crystal lattices it is also necessary to take into account the possibility of the existence of a new kind of symmetry element consisting of a combination of rotations or reflections with translations. Such elements are called *screw axes* and *glide planes*.

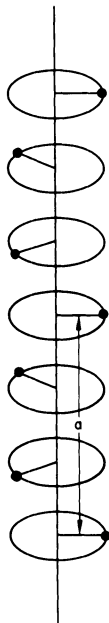


FIG. 61.

The lattice has a screw axis of order n if it is unchanged in appearance by rotation through an angle $2\pi/n$ about the axis together with a displacement through a certain distance along the axis. To illustrate this symmetry, Fig. 61 shows a linear sequence of atoms (to be imagined extended indefinitely in both directions) having a screw axis of order 3. This structure is periodic, with period a ; it is unchanged in appearance by a rotation through 120° about the axis together with a displacement through $\frac{1}{3}a$ along the axis.

If the lattice is unchanged in appearance by reflection in a plane together with a displacement through a certain distance in a direction lying in that plane the lattice is said to possess a glide plane.

Thus an actual crystal has a certain translational symmetry (described by the type of its Bravais lattice) and may also have

simple and screw axes of symmetry, rotary-reflection axes, and simple and glide planes of symmetry. All these elements may be combined in various ways.

The set of all symmetry elements of an actual crystal lattice is called its *space group*. This gives the most complete description of the symmetry of the arrangement of atoms in the crystal, i.e. the symmetry of its internal structure.

There are found to be altogether 230 different space groups, discovered by E. S. Fedorov in 1891. These groups are customarily assigned to the crystal systems in accordance with the Bravais lattices by which they are generated. We shall not, of course, list here all the space groups, but merely state how they are distributed among the systems: triclinic 2, monoclinic 13, orthorhombic 59, rhombohedral 7, tetragonal 68, hexagonal 45, cubic 36.

The phenomenon of stereoisomerism in molecules has been described in §41. This can occur also in crystals (where it is called *enantiomorphism*). There exist crystals whose lattices are mirror images and which nevertheless cannot be made to coincide by any displacement in space. As with molecules, enantiomorphism of crystals is possible only when the crystal lattice has no element of symmetry which includes reflection in a plane. An example of such a structure is given by crystals of ordinary quartz, which belongs to the rhombohedral system (this refers to the modification of quartz which exists at ordinary temperatures).

§45. Crystal classes

There are many physical phenomena in which the atomic structure of matter does not appear directly. In considering such phenomena, matter may be regarded as a continuous medium and its internal structure may be ignored. For example, the thermal expansion of solids and the deformation of solids by external forces are phenomena of this type. The properties of matter as a continuous medium are called *macroscopic properties*.

The macroscopic properties of a crystal are different in different directions. For example, the properties of transmission of light through a crystal depend on the direction of the ray; the thermal expansion of a crystal is in general different in different directions; the deformation of a crystal depends on the orientation of the external forces, and so on. The cause of this dependence of properties on direction is related, of course, to the structure of the

crystal. It is clear, for example, that the stretching of a cubic crystal in a direction parallel to the edges of the cubic cells in its lattice will not occur in the same way as a stretching along the diagonal of these cells.

The dependence of the physical properties of a body on direction is called *anisotropy*. We may say that a crystal is an anisotropic medium. In this respect crystals are fundamentally different from *isotropic* media, such as liquids and gases, whose properties are the same in all directions.



FIG. 62.

Although the properties of crystals are in general different in different directions, there may be some directions in which they are the same; such directions are said to be equivalent. For example, if a crystal has a centre of symmetry, any direction in it is equivalent to the opposite direction; if a crystal has a plane of symmetry, any direction is equivalent to the direction which is its mirror image in the plane (Fig. 62), and so on.

It is evident that the "symmetry of directions" in a crystal, and therefore the symmetry of its macroscopic properties, are determined by its axes and planes of symmetry. The translational symmetry is here unimportant, since a translation of the lattice does not affect directions in it; thus the macroscopic properties of a crystal do not depend on its particular Bravais lattice (among those possible in a given system). From this point of view it is also immaterial whether the crystal has a simple or a screw axis of symmetry of a given order, and whether a plane of symmetry is a simple or a glide plane.

There is a limited number (32) of possible combinations of planes and axes of symmetry which can describe the symmetry of directions in a crystal. These combinations, i.e. types of macroscopic symmetry of a crystal as an anisotropic medium, are called *crystal classes*.

The relation between the space group and the class of a crystal is clear from the foregoing. The class is derived from the space group by omitting all translations and the distinctions between simple and screw axes of symmetry and simple and glide planes of symmetry.

The crystal classes, like the space groups, are assigned to the systems in accordance with the Bravais lattices for which they actually occur in crystals. It is found that the numbers of classes in the systems are: triclinic 2, monoclinic 3, orthorhombic 3, tetragonal 7, cubic 5, rhombohedral 5 and hexagonal 7 (though it should be noted that all the classes in the rhombohedral system can be given by either a rhombohedral or a hexagonal Bravais lattice).

Among the classes belonging to a given system there is one which has the full symmetry of the system. The remaining classes are of lower symmetry, i.e. have fewer symmetry elements than the system in question.

As an example of the relation between the macroscopic properties and the symmetry of a crystal, let us consider thermal expansion.

An isotropic body (a liquid or gas) expands uniformly in all directions on heating, and is therefore described by a single coefficient of thermal expansion. It is easily seen that the same is true of cubic crystals. For a crystal of the cubic system, as it expands, must remain a cubic crystal, and its lattice must therefore retain its shape; hence it follows that such a crystal must expand uniformly in all directions, i.e. like an isotropic body.

A tetragonal crystal, on the other hand, though it remains tetragonal, need not retain the same ratio of the height c and width a of its cells. The crystal can therefore expand differently in the direction of the height of the cells and in directions perpendicular thereto. In other words, the thermal expansion of a tetragonal crystal is described by two coefficients (and the same is true of any uniaxial crystal). The thermal expansion of biaxial crystals is described by three coefficients which give the expansion along three axes.

§46. Lattices of the chemical elements

We shall now describe the structure of some actual crystals, and first mention that, although we speak for brevity of the atoms' being located at the lattice points, it would be more correct to say

that their nuclei are located there. The atoms themselves can not be regarded as points in a crystal lattice; they occupy a considerable volume, and neighbouring atoms are as if they were in contact. Here, as in molecules, the outer parts of the electron shells are appreciably distorted and "shared", in comparison with the shells of isolated atoms. The most accurate and complete way of describing the structure of a crystal therefore consists in determining the distribution of the "electron density" throughout the volume of the lattice.

Let us first consider the crystal structure of the chemical elements. About forty different lattices formed by the elements are known, and some of them are very complex. For example, one modification of manganese crystallises with a body-centred cubic Bravais lattice containing fifty-eight atoms in one cubic cell (29 atoms in the unit cell); one modification of sulphur has a face-centred orthorhombic Bravais lattice with 128 atoms in one cell (32 atoms in the unit cell). The great majority of the elements, however, crystallise in comparatively simple lattices.

About twenty elements form cubic crystals in which all the atoms constitute a single face-centred Bravais lattice; they include many metals (silver, gold, copper, aluminium, etc.) and also the crystals of the noble gases. In the crystals of about fifteen elements, all of which are metals, the atoms constitute a single body-centred cubic Bravais lattice; these include the alkali metals lithium, sodium and potassium. No element, however, forms a simple cubic lattice.

In order to understand the reason for this preference for body-centred and face-centred structures, let us consider a problem of a kindred type, though it has no direct physical significance: the packing of similar spheres.

Let us take first the packing of spheres in a simple cubic lattice. Then the spheres at adjoining vertices of cubic cells are in contact, and the edge a of the cube is therefore equal to the diameter d of the spheres. Since each cubic cell in this lattice corresponds to one sphere, we can say that the volume per sphere is $a^3 = d^3$. The volume of the sphere itself is $(4\pi/3)d^3/8 = 0.52d^3$, i.e. is only 52% of the volume of the cell.

A closer packing is given by the body-centred cubic lattice. Here the nearest neighbours, which must be in contact, are the atoms at a vertex and at the centre of a cell. Since the spatial

diagonal of the cube is $a\sqrt{3}$, we must have $d = a\sqrt{3}/2$, and the volume of the cubic cell is therefore $a^3 = 8d^3/3\sqrt{3}$. The body-centred cubic cell contains two spheres, and the volume of the unit cell containing one sphere is $4d^3/3\sqrt{3}$. Thus we easily find that the sphere occupies 68% of this volume.

Finally, the closest packing is that given by the face-centred cubic lattice (for which reason it is called *cubic close packing*). In this case a sphere whose centre is at the centre of a face must touch spheres whose centres are at vertices of the cube. The cube edge length $a = d\sqrt{2}$, the volume of the unit cell is one quarter of the cube volume, $\frac{1}{4}a^3 = d^3/\sqrt{2}$, and the sphere occupies 73% of this.

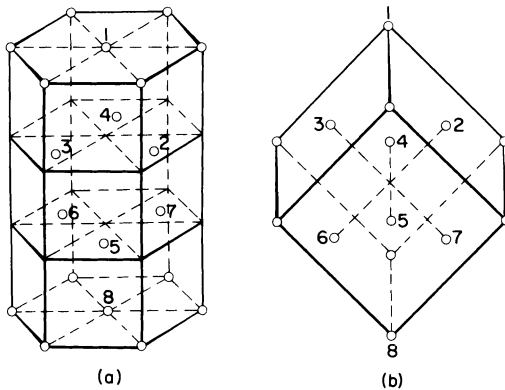


FIG. 63.

If this lattice is viewed along a diagonal of the cube, it can be seen to consist of successive layers in each of which the lattice points (sphere centres) form a network of equilateral triangles (Fig. 63a). In each successive layer the lattice points lie above the centres of the triangles in the previous layer, and there are three types of layer which alternate in succession. In Figs. 63a and 63b the figures indicate the correspondence between the points in these layers and the points of the cubic lattice.

An equally close packing may, however, clearly be achieved by alternating layers of only two types (Fig. 64). This gives a hexagonal lattice with two atoms in the unit cell, called *hexagonal close packing*. In the sphere model the ratio of the height c of the prismatic cell in this lattice (distance between nearest

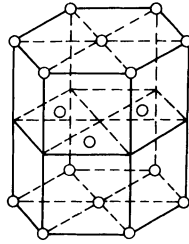


FIG. 64.

similar layers) to the length a of its base edge can be shown by calculation to be $c/a = 1.63$.

Some fifteen elements, all of which are metals, have a hexagonal close-packed lattice; they include magnesium, cadmium, zinc and nickel. In most of these the ratio of axes in the crystal is very close to the ideal value 1.63. There are exceptions, however: in cadmium and zinc the ratio c/a is about 1.9, i.e. the lattice is more elongated in the direction of the prism altitude than would occur in close packing of spheres. This results in a more marked anisotropy of these crystals.

The three types of lattice described above are those most commonly found among the elements. There exist also various other lattices in which very few elements crystallise. Some of these will be described in outline below.

The most common modification of carbon, namely graphite, has a hexagonal lattice; no other element crystallises in this form. It has a layer structure, consisting of plane parallel layers with the atoms at the vertices of regular hexagons (Fig. 65). The distance

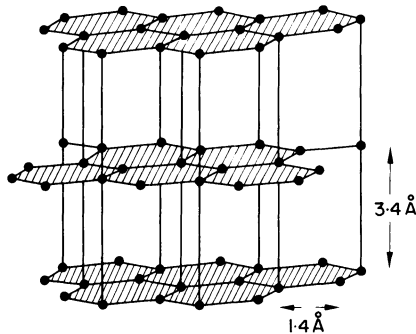


FIG. 65.

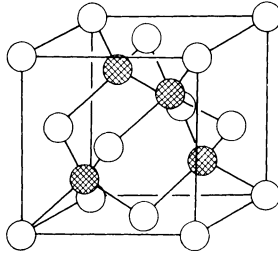


FIG. 66.

between adjoining layers is 2·3 times the distance between atoms within a layer. This explains the easy flaking of graphite.

Another modification of carbon, diamond, has a cubic lattice which may be regarded as formed by two face-centred Bravais lattices, a quarter-diagonal of the cube apart. Thus each carbon atom is surrounded by four neighbours at equal distances, forming a tetrahedron. This lattice is shown in Fig. 66; the hatched circles and the white circles are carbon atoms forming the different Bravais lattices. Silicon and germanium, the homologues of carbon, also have lattices of the diamond type.

The bismuth lattice is of interest. It belongs to the rhombohedral system, but is distinctive in being very nearly cubic. This lattice may be regarded as a slightly deformed simple cubic lattice, the cube being slightly flattened along its diagonal (thus becoming a rhombohedron), and there is also a very slight additional shift of the atoms.

All the elements described above have *atomic* lattices, in which separate molecules cannot be distinguished. Some elements, however, crystallise in *molecular lattices*. For example, hydrogen, nitrogen, oxygen and the halogens (fluorine, chlorine, bromine and iodine) form lattices which can be regarded as being composed of diatomic molecules, i.e. pairs of atoms much closer together than the distances between pairs.

§47. Lattices of compounds

The crystal lattices of chemical compounds are almost as various as the compounds themselves. Here we shall describe only some of the simplest lattices.

One of the commonest structures is that of rock salt, NaCl: a cubic lattice with half the lattice points occupied by sodium atoms

and half by chlorine atoms (Fig. 67). Each sodium atom is symmetrically surrounded by six chlorine atoms, and *vice versa*. The Bravais lattice of NaCl is a face-centred cubic lattice. Each unit cell contains two atoms, one of sodium and one of chlorine.

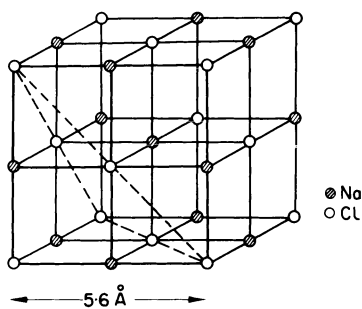


FIG. 67.

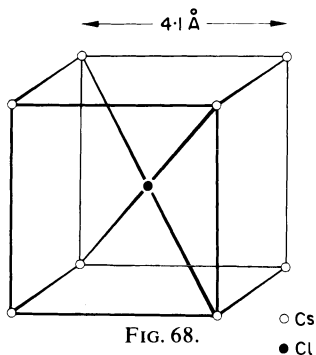


FIG. 68.

The arrangement of the atoms in a crystal lattice is customarily described by stating their coordinates (defined as in §43). It is sufficient to indicate the positions of the minimum number of atoms from which those of the remainder can be obtained by adding a lattice vector. For example, the structure of NaCl is described by the coordinates of two atoms relative to the axes of the cubic cell: Na(0, 0, 0), Cl($\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$). The coordinates of all other atoms are obtained from these by adding (or subtracting) a certain number of basic lattice vectors, which may be taken, for example, as the distances from the origin to the centres of the three faces of the cube (the points with coordinates (0, $\frac{1}{2}$, $\frac{1}{2}$), ($\frac{1}{2}$, 0, $\frac{1}{2}$), ($\frac{1}{2}$, $\frac{1}{2}$, 0)).

The lattice of caesium chloride, CsCl, is also of a very common type (Fig. 68). It has a simple cubic Bravais lattice. Atoms of one kind are at the vertices of the cubic cells, and atoms of the other kind are at the centres of the cells.

We may also mention the lattice of zinc blende, ZnS. This is obtained from the diamond lattice described in §46 by placing different atoms (Zn and S) at the points of the two interpenetrating face-centred Bravais lattices (the hatched and white circles in Fig. 66). Each zinc atom is surrounded by four sulphur atoms at the vertices of a tetrahedron, and *vice versa*. The positions of the atoms in the cubic cell are given by the coordinates Zn(0, 0, 0), S($\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$).

A characteristic property of these lattices is that molecules of the compounds cannot be distinguished as particular groups of atoms. The whole crystal is, as it were, one huge molecule.

The distribution of electrons in these lattices is such that around some nuclei there are on average more electrons, and around others fewer, than would be present in the free neutral atom. Such lattices may be quite adequately described as consisting of ions, and are therefore called *ionic* lattices. For example, the NaCl lattice consists of positive ions Na^+ and negative ions Cl^- .

There are other lattices of compounds in which individual molecules can be distinguished as especially closely arranged groups of atoms; these include, in particular, many organic crystals. But the division of crystals into atomic and molecular is largely arbitrary, and various intermediate cases are possible.

A typical example of this is the CdI_2 lattice, which has a kind of layered structure. On either side of each layer of cadmium atoms and close to it, there is a layer of iodine atoms, the distance between such "triple" layers being greater. Although the latter suggests a molecular composition of the substance, it is not possible to distinguish individual molecules within each layer.

§48. Crystal planes

In the study of crystals it is frequently necessary to consider various planes passing through the lattice. These may be planes forming natural faces of the crystal, or planes having certain physical properties: for example, if a crystal is cleaved with a knife, the cleavage usually occurs along particular planes having distinctive properties. Finally, a consideration of various planes in the lattice is necessary in structural analysis by the use of X-rays.

It is clear that only planes which pass through atoms in the crystal (i.e. through its lattice points) can have particular physical properties. Such planes are called *crystal planes*, and it is these which we shall now discuss.

It has already been mentioned in §43 that, in describing crystals, use is made of a coordinate system (in general oblique) whose axes are related in a definite way to the edges of the Bravais lattice cell, the coordinates being measured in terms of the lengths a , b , c of these edges (which are in general different). Let

these coordinates be denoted by x, y, z . The coordinates of the Bravais lattice points are given by integers (or half-integers, but this does not affect the subsequent discussion).

The general equation of a plane is

$$lx + my + nz = k$$

in either rectangular or oblique coordinates. If l, m, n and k are integers, this equation, regarded as a single equation for the three unknowns x, y, z , has an infinity of integral solutions. In other words, the plane contains an infinity of lattice points, and is therefore a crystal plane.

The significance of the numbers l, m, n is easily seen. Putting $y = z = 0$ in the equation, we find $x = k/l$; this is the coordinate of the point where the plane intersects the x axis. Similarly, we find that the intercepts of the plane on the y and z axes are k/m and k/n . Hence we conclude that the lengths of the intercepts on the three axes are in the ratios

$$1/l : 1/m : 1/n,$$

i.e. are inversely proportional to the numbers l, m, n . These lengths are measured in terms of a, b, c ; in ordinary units the lengths are in the ratios

$$a/l : b/m : c/n.$$

Thus we see that the numbers l, m, n determine the direction of the plane, i.e. its orientation relative to the axes of the lattice; the number k depends not on the direction of the plane but on its

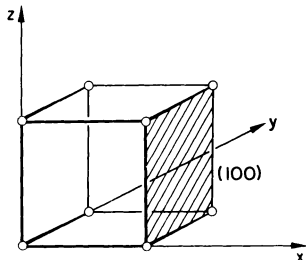


FIG. 69.

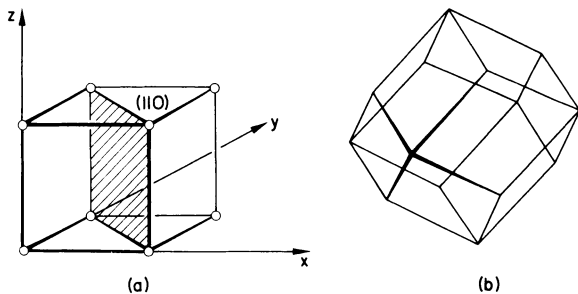


FIG. 70.

distance from the origin. By giving k various integral values, with fixed values of l, m, n , we obtain a family of parallel crystal planes. It is the direction of a crystal plane which is of importance, and not its absolute position in the lattice. In this sense the plane is fully defined by the set of three numbers l, m, n . The highest common factor may also be cancelled from these numbers, since the direction of the plane is obviously unchanged by this. The numbers l, m, n thus defined are called the *indices* of the crystal plane and are written in parentheses: (lmn) .

As examples, we shall consider various planes in a cubic lattice. The plane perpendicular to the x axis (Fig. 69) has intercepts $1, \infty, \infty$ on the axes; the reciprocals of these are $1, 0, 0$, and the indices of the plane are (100) . Similarly, the indices of the planes perpendicular to the y and z axes are (010) and (001) . These planes bound a body of cubical shape and are therefore often called *cube planes*.

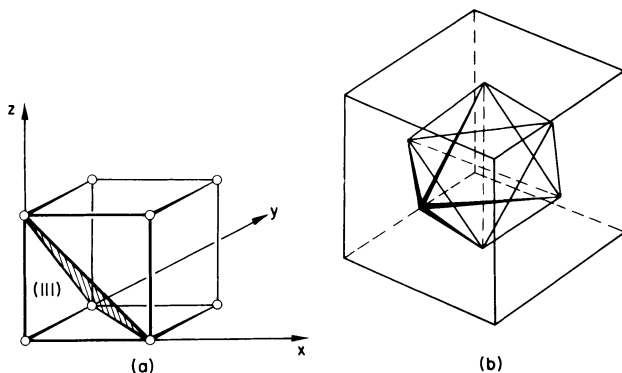


FIG. 71.

A diagonal plane parallel to the z axis has equal intercepts on the x and y axes (Fig. 70a). Its indices are therefore (110). Such diagonal planes are called *rhombic dodecahedron planes*, from the name of the dodecahedron bounded by planes of this kind (Fig. 70b).

A diagonal plane of the cube (Fig. 71a) has equal intercepts on all three axes, and its indices are therefore (111). Planes of this kind are called *octahedron planes*, from the regular octahedron with triangular faces which they form; the octahedron shown in Fig. 71b is obtained by joining the centres of the six faces of the cube.

§49. The natural boundary of a crystal

The planes which form the boundaries of a natural crystal always pass through atoms in its lattice, and are therefore crystal planes. The directions of the various faces of the crystal and the angles between them are related to the structure of its lattice and are therefore characteristic of any given substance.

Let us consider any two faces of the crystal, with indices (lmn) and ($l' m' n'$). We denote by A, B, C and A', B', C' the intercepts of these planes on the coordinate axes. According to the discussion in §48, the ratios of these intercepts (measured in ordinary units) are

$$A : B : C = a/l : b/m : c/n, \quad A' : B' : C' = a/l' : b/m' : c/n'.$$

Dividing these ratios, we obtain

$$A/A' : B/B' : C/C' = l'/l : m'/m : n'/n.$$

On multiplying by the least common multiple of l, m, n , the right-hand side of this relation is converted to the ratios of three integers.

Thus we see that the ratios of the intercepts of any face of the crystal on the axes, when expressed in terms of the intercepts of any other face, are always ratios of integers. This is called the *law of rational indices*.

The surfaces of ionic crystals must necessarily contain ions of different signs. Crystal planes containing ions of one sign only can not be crystal faces. This often provides an explanation of certain properties of the crystallisation of various substances.

Let us consider, for example, the NaCl crystal, whose lattice is shown in Fig. 67 (§47). The diagram shows how the Na^+ and Cl^- ions are situated in the (100) and (111) planes of this lattice. We see that the (111) plane (the diagonal plane shown by the broken lines in Fig. 67) passes through ions of one kind only, and this plane therefore cannot be a crystal face. Thus rock salt cannot crystallise in octahedra. The (001) plane (the cube face in Fig. 67), however, contains ions of opposite sign alternating in both directions; thus rock salt can crystallise in cubes.

In the caesium-chloride lattice (Fig. 68, §47), on the other hand, the (100) planes contain ions of only one sign, and this substance therefore can not crystallise in cubes.

The nature of the external boundary of a crystal, like all its macroscopic properties, depends on the crystal class. Thus a study of the shape of natural crystals enables us, in principle, to determine their symmetry classes. In practice this may be rendered difficult by irregularities of shape due to various accidental effects of the conditions in which the crystal was grown. Further information may be obtainable from the artificial formation of new faces by etching the surface of the crystal with a solvent.

CHAPTER VII

HEAT

§50. Temperature

In all bodies existing in Nature there is a continual movement of their constituent particles. This movement is universal: the molecules are moving, and so are the atoms within them. The characteristic feature of this movement is the randomness which it always to some extent possesses. The movement is called *thermal motion*, and it is the underlying cause of the phenomena of heat.

Although usually the term “thermal motion” refers to the motion which takes place on an atomic or *microscopic* scale, thermal motion is also a property of larger, macroscopic, particles. A well-known example of this is the *Brownian motion*, the random motion of fine particles suspended in a liquid, which may be observed through a microscope.

If two bodies are brought into contact, the atoms in them will collide and transfer energy to one another. Thus, when two bodies are in contact, energy passes from one body to the other; the body which loses energy in this process is said to be the hotter, and that which gains energy the colder body. This transfer of energy continues until a definite state of *thermal equilibrium* is set up.

To describe the hotness of bodies, the concept of *temperature* is used. A quantitative definition of this might in principle be given by using any property of bodies which depends on their hotness. For example, we could define a scale of temperature simply by the volume of a column of mercury in thermal equilibrium with the body concerned. It is evident, however, that such a temperature scale would be entirely arbitrary and would have no particular physical significance; the temperature thus defined would be extremely inconvenient for the quantitative description of any other thermal phenomena. It is therefore necessary to establish first of all a temperature scale having a physical signifi-

cance and not dependent on the nature of any one material, such as mercury or the glass of the vessel which contains the mercury.

In physics the *thermodynamic* or *absolute scale* of temperature is used; it is intimately related to the general thermal properties of all bodies. It cannot be precisely defined here, since this would require a theoretical analysis of thermal phenomena which is outside the scope of this book. Instead, we shall describe the scale by means of some of its "secondary" properties.

It is clear that a physical definition of temperature must be based on a physical quantity which describes the state of a body and which is necessarily the same for any two bodies in thermal equilibrium. The mean kinetic energy of the translational motion of the particles (molecules or atoms) in a body is in fact found to have this remarkable property. If the mean energies are the same for the particles in any two bodies, then, when the bodies are brought into contact, individual particles will transfer energy in both directions, but there will be no net transfer of energy from either body to the other.

For this reason, the mean kinetic energy of the translational motion of the particles within the body may be taken as a measure of temperature. The temperature T is customarily defined as two-thirds of this energy:

$$T = \frac{2}{3} \cdot \frac{1}{2} \overline{mv^2} = \frac{1}{3} \overline{mv^2}.$$

Here m is the mass and v the velocity of a particle, and the bar over an expression denotes that its mean value is to be taken. (The mean value may be understood as the mean energy of various particles in the body at a given instant, or as the mean energy of a given particle at various instants, the two definitions being entirely equivalent.)

According to the above definition, temperature has the dimensions of energy, and may therefore be measured in the same units as energy, such as ergs. The erg, however, is an extremely inconvenient unit for the measurement of temperature, mainly because the energy of the thermal motion of particles is usually very small in comparison with the erg. Moreover, the direct measurement of temperature as the energy of particles would of course be very difficult to carry out in practice.

For these reasons, a conventional but convenient unit of temperature, the *degree*, is used in physics. It is defined as one-hundredth of the difference between the boiling point and the freezing point of pure water at atmospheric pressure.

The conversion factor which determines the degree as a fraction of the erg is called *Boltzmann's constant*, and is usually denoted by k . Its value is

$$k = 1.38 \times 10^{-16} \text{ erg/deg.}$$

We see that the degree is in fact very small compared with the erg. As a further illustration we may find the change in the total kinetic energy of the particles in one gram-molecule of matter which corresponds to each degree of temperature change. This is obtained by multiplying k by Avogadro's number N_0 :

$$\begin{aligned} kN_0 &= 1.38 \times 10^{-16} \times 6.02 \times 10^{23} \text{ erg} \\ &= 8.31 \text{ J.} \end{aligned}$$

We may also give the conversion factor between the degree and the electron-volt, the latter being the unit of energy generally used in atomic physics:

$$\begin{aligned} 1 \text{ eV} &= 1.60 \times 10^{-12} \text{ erg} \\ &= \frac{1.60 \times 10^{-12}}{1.38 \times 10^{-16}} \text{ deg} = 11\,600 \text{ deg.} \end{aligned}$$

In what follows we shall always denote by T the temperature measured in degrees. The temperature measured in ergs is then kT , and so its definition given above must be written

$$kT = \frac{1}{3} \overline{mv^2}.$$

Since the kinetic energy is positive, so is the temperature T . It should be emphasised that this property of the temperature is not to be regarded as a law of Nature; it is simply a consequence of the definition of temperature.

As already mentioned, the scale of temperature thus defined is called the absolute scale. The zero of temperature on this scale is

the temperature at which the thermal motion ceases entirely. The scale of absolute temperature measured from this *absolute zero* is called the *Kelvin scale*, and degrees on it are denoted by the symbol °K.

Besides the Kelvin scale, another scale is widely used in practice, in which the temperature is measured from the freezing point of water, arbitrarily taken as the zero of temperature. This is called the *Celsius scale*, and degrees on it are denoted by the symbol °C.

To convert temperatures from one scale to the other, it is necessary to know the absolute temperature of the freezing point of water. According to recent measurements this is 273.15°K . Correspondingly, on the Celsius scale absolute zero is -273.15°C .

In the following, T will always denote the absolute temperature; the Celsius temperature, if needed, will be denoted by t . Clearly $T = t + 273.15^\circ$.

An experiment is often said to be conducted at room temperature, meaning 20°C (i.e. about 293°K). It is useful to note that this temperature measured in electron-volts is about $1/40$ eV.

To describe the velocity of the thermal motion of particles, we can use the square root of the quantity \bar{v}^2 which appears in the definition of temperature; this square root is usually called simply the *thermal velocity* and denoted by v_T :

$$v_T = \sqrt{\bar{v}^2} = \sqrt{(3kT/m)}.$$

This formula determines the thermal velocity of an atom, a molecule or a Brownian particle, according to the mass that is substituted in it. When the application is to molecules, it is convenient to modify the formula somewhat by multiplying the numerator and denominator of the radicand by Avogadro's number and using the fact that the product mN_0 is the molecular weight μ of the substance:

$$\begin{aligned} v_T &= \sqrt{(3N_0kT/\mu)} \\ &= 15.8 \times 10^3 \sqrt{(T/\mu)} \text{ cm/sec.} \end{aligned}$$

For example, the thermal velocity of molecules of hydrogen (H_2 , $\mu = 2$) at room temperature, is 1.9×10^5 cm/sec, i.e. about 2 km/sec.

We see that the thermal velocity is proportional to the square root of the temperature and inversely proportional to the square root of the mass of the particles. The latter relation is the reason why the thermal motion, which is very violent for molecules, is still appreciable for the microscopically small particles in the Brownian motion, but entirely negligible for massive bodies.

Let us return to the definition of temperature given above. It must be emphasised that this definition is based on classical mechanics. The quantitative relation which it asserts between the temperature and the energy of the thermal motion of the particles is valid only so long as this motion can be described by classical mechanics. It is found that, as the temperature decreases and the particle energy diminishes, the conditions for classical mechanics to be valid are eventually no longer satisfied, and classical mechanics must be replaced by quantum mechanics. This occurs sooner for particles of smaller mass and for those whose motion is more restricted by the forces acting. For example, the molecules of gas in translational motion move almost as free particles, and this motion can always be treated classically, but the motion of atoms in the molecule is of the nature of small oscillations in a "potential well" around certain equilibrium positions, and classical mechanics very soon ceases to be applicable to this motion. We shall return to this subject in §§57 and 58.

It has been mentioned above that the thermal motion no longer occurs at absolute zero. This does not mean, however, that all motion of the particles in a body has stopped. According to quantum mechanics, the motion of the particles never ceases completely. Even at absolute zero there must remain some vibrational motion of the atoms within molecules, or vibrations of the atoms about the crystal lattice points in a solid. This motion, called *zero-point vibrations*, is a quantum phenomenon, and its energy is a measure of the quantum nature of a given object. A comparison of the energy of the thermal motion of the particles with the energy of their zero-point motion may serve as a criterion of the applicability of classical mechanics: the latter is suitable for the description of the thermal motion of the particles if the energy of this motion is sufficiently large compared with the zero-point energy.

The most striking instance of the zero-point motion, which is fully maintained even at absolute zero, is the motion of electrons,

the lightest particles, in atoms. The motion of electrons within the atom is always a purely quantum phenomenon. Owing to its relatively high energy, it is affected only to a very slight extent by the temperature of the body. The thermal motion of the atoms has a considerable effect on their electron shells only at very high temperatures, of the order of many thousands of degrees.

§51. Pressure

On account of the thermal motion of its particles, a gas (or liquid) exerts a pressure on the walls of the vessel containing it. The gas molecules, on colliding with the walls, transfer some momentum to them, and the change in momentum of a body per unit time defines the force acting on it.

The force exerted by the gas (or liquid) per unit area of the wall gives the *pressure* on the wall of the vessel, which will be denoted by p . The dimensions of pressure are those of force divided by those of area, and can be written in various ways:

$$[p] = \frac{\text{dyn}}{\text{cm}^2} = \frac{\text{erg}}{\text{cm}^3} = \frac{\text{g}}{\text{cm} \cdot \text{sec}^2}.$$

In particular, it should be noted that the dimensions of pressure are the same as those of energy per unit volume.

The unit of pressure in the CGS system is 1 dyn/cm^2 : a force of one dyne acting on an area of one square centimetre. This unit is very small, however. A unit 10^6 times larger is called a *bar*:

$$1 \text{ bar} = 10^6 \text{ dyn/cm}^2 = 10 \text{ N/m}^2.$$

The pressure at which a force of 1 kgf acts on an area of 1 cm^2 is called a *metric or technical atmosphere* (at):

$$1 \text{ at} = 1 \text{ kgf/cm}^2 = 0.981 \text{ bar}.$$

The *standard atmosphere* (atm) is the pressure of a column of mercury of height 760 mm (with a certain density of mercury and a standard acceleration due to gravity):

$$1 \text{ atm} = 1.013 \text{ bar} = 1.033 \text{ at}.$$

The pressure corresponding to one millimetre of mercury is

$$1 \text{ mm Hg} = 1.333 \times 10^{-3} \text{ bar.}$$

The properties of bodies taken as a whole without considering the details of their molecular structure (on which these properties in fact depend) are called *macroscopic* properties. Temperature and pressure are among the most important quantities describing the macroscopic state of a body; another such quantity is the volume of the body (denoted by V). These three quantities, however, are not independent. For example, if a certain quantity of gas is enclosed in a vessel of given volume and has a given temperature, its pressure is thereby determined; if the volume or the temperature is changed, the pressure of the gas also changes.

Thus only two of the three quantities p , V , T can be arbitrarily specified, the third being a function of these. We may say that the thermal properties are entirely determined by specifying any two of these quantities.

The functional relation between the pressure, volume and temperature of a body is called the *equation of state* of the body concerned, and is one of the most important relations describing its thermal properties.

The theoretical form of this relation can be established only for the simplest substances (see §53). In practice, therefore, experimental measurements must be used, the results of which can be represented graphically. Since a relation between three quantities is concerned, it would be fully represented by a surface in a three-dimensional coordinate system with p , V and T plotted along the axes. However, since a three-dimensional construction is inconvenient in practice, only two-dimensional diagrams are generally drawn, showing families of curves which are the intersections of the surface with various planes parallel to one of the coordinate planes. For example, by taking the intersections of the surface with planes parallel to the pV plane, i.e. perpendicular to the T axis, we obtain a family of curves called *isotherms*, which give the pressure as a function of the volume of the body for various given values of the temperature. Similarly *isobars* can be drawn; these are curves which give V as a function of T for given values of p . Finally, *isochores* give p as a function of T for given values of V .

It has already been mentioned in §50 that the exchange of energy between bodies in contact continues until their temperatures are equal and thermal equilibrium is reached. A state of *thermal equilibrium* of a system of bodies is defined as a state in which no spontaneous thermal processes occur in the system and every part of the system is at rest relative to the other parts and has no macroscopic motion (as opposed to the microscopic thermal motion of the particles within bodies). We may now add that in equilibrium not only the temperatures but also the pressures of all bodies in contact must be equal, since otherwise the total forces on the bodies would not be zero and the bodies would begin to move.

Under ordinary conditions the pressure is positive, i.e. is in the same direction as if the body were tending to expand. This is not necessary, however, and states of negative pressure are also possible, in which the body is as if it were “stretched” and therefore tends to contract. “Stretched” states of a liquid can be brought about by sealing a carefully purified heated liquid in a thick-walled capillary. When the capillary cools, if its walls contract more slowly than the liquid, the latter should occupy only part of the volume within the capillary. The liquid, however, adheres to the walls and is thus “stretched” over the whole volume of the capillary. Another method is to place a liquid in a glass capillary open at each end, which is then rapidly rotated about its midpoint. The liquid is stretched by centrifugal forces and, when a certain speed of rotation is reached, it finally “breaks” and is thrown out of the capillary. Considerable negative pressures can be attained by these methods: up to 280 atm in water (at room temperature), up to 40 atm in alcohol, up to 160 atm in benzene, and so on. These values may be regarded as representing the resistance of the liquid to disruption.

§52. Aggregate states of matter

The concept of *aggregate states* (gaseous, liquid and solid) is used to give the most general description of the thermal properties of bodies.

Owing to the low density of matter in the gaseous state, its molecules are relatively far apart, being at distances large compared with the size of the molecules themselves. The interaction between the molecules of a gas is therefore of subordinate importance, and for the greater part of the time the molecules move

freely, undergoing collisions with one another only quite rarely. In liquids, on the other hand, the molecules are at distances comparable with their own dimensions, so that they are all in continual strong interaction and their thermal motion is highly complicated and irregular.

Under ordinary conditions, liquids and gases differ so greatly in density that there is no difficulty in distinguishing between them. Nevertheless, the difference between these two states of matter is in fact not fundamental, but merely quantitative, arising from the value of the density and the consequent degree of interaction of the molecules. The lack of any fundamental distinction between them is especially clear from the fact that the transition between a liquid state and a gaseous state can, in principle, be completely continuous, so that there is no instant at which we can say that one state ceases to exist and the other commences. This will be discussed further in §69.

The difference between liquids and what are called *amorphous* (non-crystalline) *solids* is also quantitative; the latter substances include glass, various resins, etc. Here again the absence of any fundamental difference is shown by the possibility of a continuous transition from one state to the other, achieved simply by heating. For example, solid glass, when heated, becomes gradually softer and finally entirely liquid; this process is completely continuous and there is no "instant of transition". The density of an amorphous solid is not greatly different from that of the liquid formed from it. The main quantitative difference between them is in viscosity, i.e. the ease with which they flow; this will be further discussed in §118.

A general property of gases, liquids and amorphous solids is that the molecules in them are randomly distributed. This brings about the isotropy of these bodies, i.e. the fact that their properties are the same in all directions. The property of isotropy makes these bodies fundamentally different from the anisotropic *crystalline solids*, in which the atoms are arranged in a regular manner.

The thermal motion of the atoms in solids consists of small oscillations about certain equilibrium positions. In crystals, these positions are the crystal lattice points; in this respect the discussion in Chapter VI was imprecise in that it referred to the lattice points as the positions of the atomic nuclei, instead of the points about which the nuclei oscillate. Although the thermal motion in

solids is more “ordered” than in gases or liquids (the atoms remaining close to the lattice points), it is random in the sense that the amplitudes and phases of the various atoms are entirely unrelated.

Almost all solid bodies are crystalline, but only rarely are they separate crystals regular throughout their volume, called *single crystals*; such crystals are formed only under special conditions of growth.

Crystalline solids usually exist as *polycrystals*; all metals, for example, are of this kind. Such bodies consist of a very large number of individual *crystallites* or *grains*, often of microscopic size; for example, the dimensions of the crystallites in metals are usually of the order of 10^{-5} to 10^{-3} cm (the size depending considerably on the methods by which the metal is produced and treated).

The relative position and orientation of the individual crystallites in a polycrystalline substance are usually entirely random. For this reason, when such a substance is considered on a scale large in comparison with the dimensions of the crystallites, it is isotropic. It is clear from the foregoing that this isotropy of polycrystalline bodies is only secondary, in contrast to their true molecular anisotropy which takes effect in the anisotropy of individual crystallites.

As a result of some particular treatment or a special method of growth, it may be possible to prepare a polycrystalline substance in which the crystallites have a preferred crystallographic orientation. Such substances are said to have *texture*. For example, texture may be produced in metals by various kinds of cold working. The properties of these substances are, of course, anisotropic.

§53. Ideal gases

The simplest thermal properties are those of a gas so rarefied that the interaction between its molecules is of no practical importance. Such a gas, in which the interactions between the molecules may be neglected, is called an *ideal gas*.

It should not be thought that the interaction between the molecules of an ideal gas does not exist at all. On the contrary, its molecules collide with one another and these collisions are important in bringing about the particular thermal properties of

the gas. But the collisions occur so rarely that the gas molecules move as free particles for the greater part of the time.

Let us derive the equation of state of an ideal gas, i.e. the relation between its pressure, volume and temperature. To do so, we imagine the gas to be enclosed in a vessel having the shape of a rectangular parallelepiped, and assume that the walls are "perfectly reflecting", i.e. reflect the incident molecules at an angle equal to the angle of incidence, without change in the magnitude of the velocity. (In Fig. 72 v and v' are the velocities of a molecule before and after the collision; they are equal in magnitude and are at the same angle α to the normal to the wall.) These assumptions are made for simplicity; it is evident that the internal properties of the gas as such cannot in fact depend either on the shape of the vessel or on the properties of its walls.

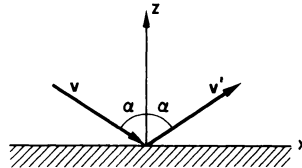


FIG. 72.

The pressure of the gas on one face of the parallelepiped may be found by determining the momentum transferred to this face per unit time by molecules colliding with it. Since only the velocity component v_z perpendicular to the surface of the wall changes in a collision, and the change is simply a change in sign, the momentum transferred in one collision is $mv_z - (-mv_z) = 2mv_z$, where m is the mass of a molecule. When moving freely, the molecule traverses the distance (h , say) between opposite walls in a time h/v_z , and so it returns to the first wall in a time $2h/v_z$. Thus each molecule has $v_z/2h$ collisions with a given wall per unit time and transfers to it a momentum $2mv_z \cdot v_z/2h = mv_z^2/h$. The total force F_z acting on the wall is the momentum transferred to it per unit time by all the gas molecules,

$$F_z = \frac{1}{h} \sum mv_z^2,$$

where Σ denotes summation over all the molecules.

If the number of molecules in the vessel is N , the sum can be written as N times the mean value $\overline{mv_z^2}$. But since all directions are completely equivalent with respect to the gas itself, we have $\overline{mv_x^2} = \overline{mv_y^2} = \overline{mv_z^2}$ and, since $v_x^2 + v_y^2 + v_z^2 = v^2$,

$$\overline{mv_z^2} = \frac{1}{3}\overline{mv^2}.$$

Thus we have

$$F_z = \frac{1}{h} \cdot \frac{N}{3} \overline{mv^2}.$$

Replacing F_z by pS , where p is the pressure of the gas and S the area of the face, and noting that hS is the volume V of the parallelepiped, we have

$$pV = \frac{1}{3}N\overline{mv^2} = \frac{2}{3}N \cdot \frac{1}{2}\overline{mv^2}.$$

The mean kinetic energy of a molecule is, from the definition of temperature, $\frac{3}{2}kT$, and thus we have finally the *equation of state for an ideal gas*:

$$pV = NkT.$$

This is a universal equation, involving no quantities dependent on the nature of the gas—a result which is an obvious effect of neglecting the interaction between the molecules and thus depriving the gas of any “individuality”.

For two different ideal gases occupying equal volumes at the same pressure and temperature, the number of molecules will be the same in each gas. This is *Avogadro's law*. In particular, one cubic centimetre of any ideal gas under standard conditions, i.e. at temperature 0°C and pressure 1 atm, contains

$$L = pV/kT = \frac{1 \cdot 013 \times 10^6 \times 1}{1 \cdot 38 \times 10^{-16} \times 273} = 2 \cdot 7 \times 10^{19} \text{ molecules};$$

this is sometimes called *Loschmidt's number*.

The number N of molecules in a given mass of gas may be written $N = \nu N_0$, where ν is the number of gram-molecules

(moles) of gas and N_0 is Avogadro's number. Then the equation of state becomes

$$pV = \nu RT,$$

where $R = kN_0$ is called the *gas constant*. In particular, for one mole of gas we have

$$pV = RT.$$

Multiplication of the values of k and N_0 gives

$$\begin{aligned} R &= 8.314 \times 10^7 \text{ erg/deg.mole} \\ &= 8.314 \text{ J/deg.mole;} \end{aligned}$$

if the calorie is used as the unit of energy, R is very nearly equal to 2 cal/deg.mole.

If the gas pressure is measured in atmospheres and the pressure in litres, then

$$R = 0.082 \text{ l.atm/deg.mole.}$$

Using this value, we can easily find the volume of one gram-molecule of gas at 1 atm pressure and 0°C :

$$V = RT/p = 0.082 \times 273/1 = 22.4 \text{ l.}$$

At constant temperature the product of the pressure and volume of a given quantity of gas is constant:

$$pV = \text{constant for } T = \text{constant.}$$

This is *Boyle's law*.

From the equation of state of an ideal gas it also follows that, if a certain mass of gas is at constant pressure, its volume is proportional to the absolute temperature of the gas:

$$V/V_0 = T/T_0 \quad \text{for } p = \text{constant,}$$

where V and V_0 are the values of the gas volume at temperatures T and T_0 . Similarly

$$p/p_0 = T/T_0 \quad \text{for } V = \text{constant.}$$

These important relations show that the absolute scale of temperature can be constructed without measuring the velocities and energies of molecules, by using the properties of ideal gases.

If T_0 is the freezing point of water, and the Celsius temperature t is used instead of the absolute temperature T of the gas ($T = 273 + t$), the above relation between the volume and the temperature of the gas may be written

$$V = V_0 \left(1 + \frac{t}{273} \right) \quad \text{for } p = \text{constant.}$$

This is *Charles' law*, according to which the volume of the gas increases by $1/273$ of its value at 0°C when the gas is heated by 1° .

In deriving the equation of state for an ideal gas we have made no use of the fact that all its molecules are identical. This equation is therefore valid also when the gas is a mixture of several different ideal gases—again a natural result of neglecting the interaction between molecules. It is only necessary to take N as the total number of gas molecules, i.e. the sum of the numbers of the various kinds of molecules: $N = N_1 + N_2 + \dots$, where N_i is the number of molecules of the i th kind. Writing the equation of state of the gas as

$$pV = N_1 kT + N_2 kT + \dots$$

and noting that, if the whole volume V were occupied by molecules of the i th kind alone, the pressure p_i would be such that $p_i V = N_i kT$, we conclude that

$$p = p_1 + p_2 + \dots,$$

i.e. the pressure of a mixture of gases is equal to the sum of the pressures which each individual gas would exert if it alone filled the volume (*Dalton's law*). The pressures p_1, p_2, \dots are called the *partial pressures* of the respective gases.

§54. An ideal gas in an external field

Let us consider an ideal gas in a force field, for example the field of gravity. Since external forces then act on the gas molecules, the gas pressure will not be the same everywhere, but will vary from point to point.

For simplicity, we shall take the case where the field forces are in a fixed direction, which we choose as the z axis. We consider two unit areas perpendicular to the z axis and at a distance dz apart. If the gas pressures on the two areas are p and $p + dp$, the pressure difference dp must clearly be equal to the total force on the gas particles in a parallelepiped of unit base and height dz . This force is $Fndz$, where n is the density of molecules (i.e. the number of molecules per unit volume) and F the force on one molecule at a point with coordinate z . Hence

$$dp = n F dz.$$

The force F is related to the potential energy $U(z)$ of a molecule by $F = -dU/dz$, so that

$$dp = -ndz \cdot dU/dz = -ndU.$$

Since the gas is assumed ideal, $pV = NkT$, and by using the relation $N/V = n$ we can write this as $p = nkT$. We shall suppose that the gas temperature is the same at every point. Then

$$dp = kTdn.$$

Equating this to the above expression $dp = -ndU$, we find

$$dn/n = d \log_e n = -dU/kT,$$

whence

$$\log_e n = -U/kT + \text{constant}$$

and finally

$$n = n_0 e^{-U/kT},$$

where n_0 is a constant which is evidently the density of molecules at a point where $U = 0$.

The formula just derived which relates the variation in density of the gas to the potential energy of its molecules is called *Boltzmann's formula*. The pressure differs from the density by a

constant factor kT , and so a similar equation is valid for the pressure:

$$p = p_0 e^{-U/kT}.$$

In the field of gravity near the Earth's surface, the potential energy of a molecule at height z is $U = mgz$, where m is the mass of a molecule. Thus, if the temperature of the gas is regarded as independent of height, the pressure p at height z is related to the pressure p_0 on the Earth's surface by

$$p = p_0 e^{-mgz/kT}.$$

This is called the *barometric formula*; it may be more conveniently written in the form

$$p = p_0 e^{-\mu gz/RT},$$

where μ is the molecular weight of the gas and R the gas constant.

This formula can also be applied to a mixture of gases. Since there is practically no interaction between the molecules of ideal gases, each gas may be treated separately, i.e. a similar formula is applicable to the partial pressure of each gas.

The greater the molecular weight of a gas, the more rapidly its pressure decreases with increasing height. The atmosphere therefore contains an increasing proportion of light gases with increasing height; the content of oxygen, for example, decreases more rapidly than that of nitrogen.

It should be remembered, however, that the applicability of the barometric formula to the real atmosphere is very limited, since the atmosphere is not in fact in thermal equilibrium and its temperature varies with height.

An interesting conclusion can be drawn from Boltzmann's formula if we attempt to apply it to the atmosphere at all distances from the Earth. At very large distances from the Earth's surface, U must be taken not as mgz but as the exact value of the potential energy of a particle:

$$U = -GMm/r,$$

where G is the gravitational constant, M the Earth's mass and

r the distance from the centre of the Earth (see §22). Substituting this energy in Boltzmann's formula gives the following expression for the gas density:

$$n = n_{\infty} e^{GMm/kTr},$$

where now n_{∞} denotes the gas density where $U = 0$ (i.e. at an infinite distance from the Earth). Putting r here equal to the Earth's radius R , we find a relation between the densities of the atmosphere at the Earth's surface (n_E) and at infinity (n_{∞}):

$$n_{\infty} = n_E e^{-GMm/kTR}.$$

According to this formula, the density of the atmosphere at an infinite distance from the Earth should be non-zero. This conclusion is absurd, however, since the atmosphere originates from the Earth, and a finite quantity of gas cannot be spread over an infinite volume with a density which is nowhere zero. The conclusion is reached because we have tacitly assumed that the atmosphere is in a state of thermal equilibrium, which does not in fact exist. This result shows, however, that a gravitational field cannot retain a gas in a state of equilibrium, and the atmosphere should therefore be steadily dissipated into space. For the Earth this dissipation is extremely slow, and in its whole time of existence the Earth has not lost an appreciable fraction of its atmosphere. For the Moon, however, with its much weaker field of gravity, the atmosphere has been lost much more quickly, and in consequence the Moon now has no atmosphere.

§55. The Maxwellian distribution

The thermal velocity v_T is a certain average property of the thermal motion of particles. In reality, different molecules move with different velocities and we may ask what is the velocity distribution of the molecules, that is, how many (on average) of the molecules in the body have a particular velocity?

We shall derive the answer to this question for an ideal gas in thermal equilibrium. To do so, let us consider a column of gas in a uniform field of gravity, and first examine the distribution of molecules with respect to the values of only one velocity component, the vertical component v_z . Let $nf(v_z)dv_z$ denote the

number of molecules per unit volume of the gas for which the value of this component lies in an infinitesimal interval between v_z and $v_z + dv_z$. Here n is the total number of molecules in the volume considered, and so $f(v_z)$ determines the fraction of molecules having a particular value of v_z .

Let us consider molecules with velocities in the interval dv_z which are in a layer of gas at height z and of infinitesimal thickness dz . The volume of this layer is equal to dz if the cross-section of the gas column is of unit area, and the number of such molecules is therefore $n(z)f(v_z)dv_zdz$, where $n(z)$ is the density of gas molecules at height z . These molecules move as free particles (since the collisions in an ideal gas may here be neglected) and subsequently reach a different layer of thickness dz' at height z' , with velocities in the interval between some values v_z' and $v_z' + dv_z'$. Since the number of molecules is unchanged, we have

$$n(z)f(v_z)dv_zdz = n(z')f(v_z')dv_z'dz'.$$

For movement in a field of gravity, the horizontal velocity components v_x, v_y remain constant, and the change in v_z is determined by the law of conservation of energy, according to which

$$\frac{1}{2}mv_z^2 + mgz = \frac{1}{2}mv_z'^2 + mgz'.$$

Differentiating this equation (for given constant values of z and z') we obtain

$$v_z dv_z = v_z' dv_z'$$

as the relation between dv_z and dv_z' , the ranges of values of the vertical velocities of the molecules considered, at heights z and z' . The thicknesses dz and dz' of the layers are related by

$$dz/v_z = dz'/v_z';$$

this simply expresses the fact that in a time $dt = dz/v_z$ during which a molecule crosses a layer dz at height z , it will travel a distance $dz' = v_z'dt$ at height z' . Multiplication of the two equations above gives

$$dv_z dz = dv_z' dz'.$$

In the condition of constant number of molecules shown previously, therefore, the differentials on the two sides of the equation cancel, leaving

$$n(z)f(v_z) = n(z')f(v_z').$$

The barometric formula states that

$$n(z)/n(z') = e^{-mgh/kT}$$

(where $h = z - z'$ is the difference in height), and hence

$$f(v_z') = f(v_z)e^{-mgh/kT}.$$

Thus the required distribution function must be multiplied by $e^{-mgh/kT}$ when $\frac{1}{2}mv_z^2$ is replaced by $\frac{1}{2}mv_z'^2 = \frac{1}{2}mv_z^2 + mgh$. The only function having this property is the exponential function

$$f(v_z) = \text{constant} \times e^{-mv_z^2/2kT}.$$

[It should be noted that the acceleration due to gravity does not appear in this formula. This is as it should be, since the mechanism of establishment of the velocity distribution of the gas molecules consists in collisions between molecules and does not depend on the external field. In the foregoing derivation the field served only the auxiliary purpose of relating the velocity distribution to the already known Boltzmann's formula.]

We have found the equilibrium distribution of molecules with respect to one component of the velocity. The fraction of molecules having given values of all three velocity components simultaneously is evidently obtained by multiplying together the fractions of molecules having given values of each component separately. Thus the complete distribution function is

$$f(v_x, v_y, v_z) = \text{constant} \times e^{-mv_x^2/2kT} e^{-mv_y^2/2kT} e^{-mv_z^2/2kT}.$$

Adding the exponents and using the fact that the sum $v_x^2 + v_y^2 + v_z^2$ is v^2 , the square of the magnitude of the velocity, we have finally

$$f = \text{constant} \times e^{-mv^2/2kT}.$$

The number dN of gas molecules whose velocity components lie in the intervals between v_x, v_y, v_z and $v_x + dv_x, v_y + dv_y, v_z + dv_z$ is therefore

$$dN = \text{constant} \times e^{-mv^2/2kT} dv_x dv_y dv_z;$$

the constant coefficient is determined by the condition that the total number of molecules with all possible velocities is equal to the given number N of molecules in the gas, but its value will not be written out here. The formula derived above is called the *Maxwellian distribution* formula.

The analogy between this formula and Boltzmann's formula for the gas density distribution in space in an external field should be noted: in each formula we have an exponential expression of the form $e^{-\epsilon/kT}$, where ϵ is the energy of a molecule (the kinetic energy $\frac{1}{2}mv^2$ for the velocity distribution, and the potential energy $U(x,y,z)$ in the external field for the distribution in space). This exponential expression is often called a *Boltzmann factor*.

If the three components v_x, v_y, v_z are given, both the magnitude and the direction of the velocity of the molecule are determined. But the distribution of molecules with respect to the direction of the velocity is simply a uniform distribution, with equal numbers, on average, travelling in every direction. [This follows from the fact that the Maxwellian distribution involves only the absolute magnitude v of the velocity, but it is also evident *a priori*: if there existed some preferred direction of motion of the molecules in the gas, this would mean that the gas was not at rest but was moving in that direction.]

The Maxwellian formula can be transformed so as to give directly the distribution of gas molecules with respect to absolute magnitude of velocity regardless of direction. For this purpose we must take the total number of molecules with various values of the velocity components v_x, v_y, v_z but a given value of $v^2 = v_x^2 + v_y^2 + v_z^2$. This is easily done by using the following geometrical analogy. If we use a coordinate system with the values of v_x, v_y, v_z plotted along the axes, the product $dv_x dv_y dv_z$ will be the volume of an infinitesimal parallelepiped with edges dv_x, dv_y, dv_z . We must sum over all volume elements at a fixed distance from the origin (since v is clearly the length of the "radius vector" in these coordinates). These volumes occupy

a spherical shell between two spheres of radii v and $v + dv$. The volume of the shell is equal to the area $4\pi v^2$ of the spherical surface multiplied by the thickness dv of the shell.

Thus, replacing the product $dv_x dv_y dv_z$ in the Maxwellian distribution formula by $4\pi v^2 dv$, we find the number of molecules with velocities in the interval from v to $v + dv$:

$$dN = \text{constant} \times e^{-mv^2/2kT} v^2 dv.$$

The coefficient of dv in this formula is the number of molecules per unit interval of velocity. As a function of v it has the form shown in Fig. 73. It is zero when $v = 0$, reaches a maximum for a value v_0 , and tends very rapidly to zero as the velocity increases further. The maximum on the curve corresponds to the value $v_0 = \sqrt{(2kT/m)}$, which is slightly less than the thermal velocity v_T defined in §50.

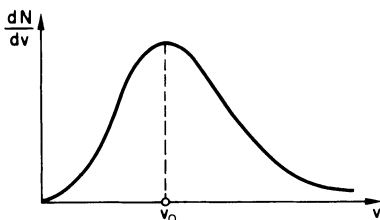


FIG. 73.

Since different molecules have different velocities, it makes a difference, in determining the mean properties, which quantity is averaged. For example, the mean value \bar{v} of the velocity itself is not the same as the velocity $v_T = \sqrt{\bar{v}^2}$ (which is often called also the root-mean-square velocity, in order to stress its origin). From the Maxwellian distribution it can be shown that $\bar{v} = 0.92v_T$.

The Maxwellian distribution has been derived here for a monatomic gas, but it can in fact be deduced from much more general theoretical arguments, and is a universal result. It is valid for the thermal motion of molecules and atoms in all bodies, but it is based on classical mechanics and its validity is limited by quantum effects in the same way as the applicability of classical mechanics in general to thermal motion.

The velocity distribution in thermal motion can be studied by various methods using *molecular beams*. These are obtained by allowing molecules to evaporate into an evacuated vessel from a substance heated in a special type of furnace. The vessel is evacuated to such an extent that molecules move in it almost without collisions.

One such method is based on the idea of a mechanical velocity selector, which works in the following way. Two circular discs with radial slots at an angle α to each other rotate on a common axis at a distance l apart in an evacuated vessel (Fig. 74). A molecular beam from the furnace F passes through the diaphragm D to the discs. A molecule which passes through the slot in the first disc with velocity v will reach the second disc after a time $t = l/v$. In this time the disc turns through an angle $\Omega t = \Omega l/v$, where Ω is the angular velocity of rotation. Thus only molecules whose velocity is such that $\Omega l/v = \alpha$ will pass through the slot in the second disc and leave a trace on the screen S . By varying the speed of rotation of the discs and measuring the density of the deposit on the screen we can find the relative numbers of particles with various velocities.

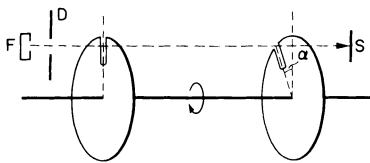


FIG. 74.

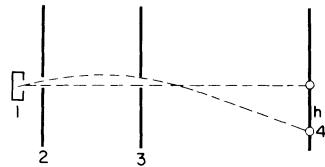


FIG. 75.

The Maxwellian distribution has also been tested experimentally by observing the deviation of a molecular beam under gravity. Atoms of caesium heated in the furnace 1 (Fig. 75) and emerging from an aperture in it enter an evacuated vessel. A narrow beam selected by the diaphragms 2 and 3 is deflected downwards by gravity and is collected by a detector in the form of a heated thin horizontal tungsten wire 4 which can be placed at various distances h below the axis of the apparatus; the caesium atoms which strike the wire leave it as positive ions which are collected by a negatively charged plate. The deflection h of an atom depends on its velocity v ; in the experiments, this deflection was some tenths of a millimetre with a beam path

length of 2 metres. By measuring the beam intensity for various values of h we can find the velocity distribution of the atoms in the beam.

§56. Work and quantity of heat

When a body expands, it moves the surrounding bodies, i.e. does work on them. Let us consider, for example, a gas beneath a piston in a cylindrical vessel. If the gas expands and moves the piston an infinitesimal distance dh , it does work dA on the piston, where $dA = F dh$ and F is the force exerted by the gas on the piston. But, by definition, $F = pS$, where p is the gas pressure and S the area of the piston. Hence $dA = pSdh$, and since Sdh is the increase dV in the volume of the gas we have finally

$$dA = p dV.$$

This simple and important formula determines the work done in an infinitesimal change in the volume of a body. We see that this work depends only on the pressure and the total change in volume, and not on the shape of the body. [To avoid misunderstanding it should be mentioned at once that this assertion does not apply to solids; see §101.]

The work dA is positive when the body expands ($dV > 0$), and the body does work on the surrounding medium. When the body is compressed ($dV < 0$), on the other hand, work is done on it by the surrounding bodies, and with our definition of dA this corresponds to negative work.

The work done in a given process can be represented by a geometrical analogy if the process is shown graphically as a curve in the coordinates p and V . For example, let the change in

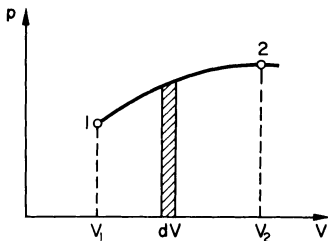


FIG. 76.

pressure of a gas as it expands be shown by the curve 12 in Fig. 76. When the volume increases by dV , the work done by the gas is $p dV$, i.e. the area of the infinitely narrow rectangle represented by the hatched area in the diagram. The total work done by the gas in expanding from volume V_1 to V_2 therefore consists of the elements of work dA whose sum is represented by the area $12V_2V_1$ below the curve and between the two extreme vertical lines. Thus the area in the diagram gives at once the work done by the body in the process considered.

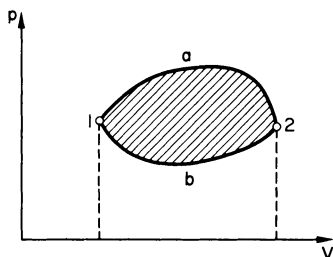


FIG. 77.

One frequently encounters *cyclic processes*, i.e. those in which the body finally returns to its original state. For example, let a gas be subjected to the process shown by the closed curve $1a2b1$ in Fig. 77. On the curve $1a2$ the gas expands and does work represented by the area under that curve; on the curve $2b1$ the gas is compressed, and the work done is therefore negative and equal in magnitude to the area under the curve $2b1$. The total work done by the gas is consequently equal to the difference of these areas, i.e. is represented by the hatched area in Fig. 77 lying within the closed curve.

The total work A done by the body in expansion from volume V_1 to V_2 is given by a particularly simple expression when the process occurs at constant pressure. In this case we clearly have

$$A = p(V_2 - V_1).$$

We may also determine the work done in an isothermal expansion of an ideal gas. For one gram-molecule of gas the pressure $p = RT/V$; hence

$$dA = p dV = (RT/V)dV = RT d \log_e V;$$

since the temperature remains constant, we can write $dA = d(RT \log_e V)$. Hence it follows that the work A is equal to the difference between the values of $RT \log_e V$ at the end and the beginning of the process, i.e.

$$A = RT \log_e (V_2/V_1).$$

If the body gains no energy from external sources, the work done in expansion is done at the expense of its internal energy. This energy, which we denote by E , includes the kinetic energy of the thermal motion of the atoms of the substance and the potential energy of their mutual interaction.

However, the change in the internal energy of the body in a given process is not in general equal to the work done. The reason is that the body may also gain (or lose) energy by direct transfer from other bodies without doing mechanical work. The energy thus gained is called the *quantity of heat* gained by the body; we shall regard it as positive if the body gains heat and negative if it loses heat.

Thus the infinitesimal change in the internal energy of the body consists of two parts: an increase due to the quantity of heat gained by the body (which we denote by dQ) and a decrease due to the work dA done by the body. Hence we have

$$dE = dQ - pdV.$$

This important relation expresses the law of conservation of energy for thermal processes and is called in this connection the *first law of thermodynamics*.

It must be emphasised that the work and the quantity of heat depend not only on the initial and final states of the body but also on the path along which the change in the state of the body takes place. For this reason we cannot speak of the “quantity of heat contained in a body” and regard the amount of heat concerned in the process as the difference of this quantity in the final and initial states. The fact that such a quantity has no meaning is especially clear if we consider a cyclic process, where the body returns to its initial state but the total amount of heat gained (or lost) is certainly not zero.

Only the internal energy E is what is called a *function of the state*: in any given state, the body has a definite energy. The

total change in the energy of a body during a process is therefore a quantity depending only on the final and initial states, namely the difference $E_2 - E_1$ between the energies in these states. The separation of this change into a quantity of heat Q and an amount of work A is not unique, but depends on the path taken in going from the initial to the final state. In particular, in a cyclic process the total change in energy is zero; the quantity of heat Q gained by the body and the work A done by it are not zero, but $Q = A$.

In thermal measurements a special unit of energy, the calorie (cal), was used until recently. The definition of this unit as the quantity of heat needed to heat 1 g of water by 1° is insufficiently exact, since the specific heat of water depends slightly on the temperature. In consequence, various definitions of the calorie existed which differed somewhat in value. The relation between the calorie and the joule is approximately

$$1 \text{ cal} = 4.18 \text{ J.}$$

If the temperature of one gram-molecule of a substance is raised by dT when it gains a quantity of heat dQ , the ratio

$$C = dQ/dT$$

is called the *specific heat* of the substance. This definition, however, is inadequate by itself, since the quantity of heat necessary depends not only on the change in temperature but also on the other conditions under which the heating takes place: it is necessary to state how other properties of the substance besides the temperature are affected. Because of this indefiniteness, various definitions of the specific heat are possible.

The most usual in physics are the *specific heat at constant volume* C_V and the *specific heat at constant pressure* C_p , which give the quantities of heat when the substance is heated under conditions such that its volume and pressure respectively remain constant.

If the volume remains constant, then $dV = 0$ and $dQ = dE$, i.e. all the heat is used to increase the internal energy. We can therefore write

$$C_V = (dE/dT)_V.$$

The suffix V to the derivative signifies that the differentiation is to be taken for a constant value of V . This indication is necessary, since the energy of a body depends, in general, not only on the temperature but also on other quantities describing the state of the body, and the result of the differentiation therefore depends on which of these quantities is assumed constant.

If the pressure remains constant on heating, then heat is used not only to increase the internal energy but also to do work. In this case the quantity of heat may be written in the form

$$dQ = dE + pdV = d(E + pV),$$

since $p = \text{constant}$. We see that the quantity of heat is equal to the change in the quantity

$$W = E + pV.$$

This is called the *enthalpy*, *heat function* or *heat content*; like the energy, it is a definite function of the state of the body. Thus the specific heat at constant pressure may be calculated as the derivative

$$C_p = (dW/dT)_p.$$

The specific heat C_p is always greater than C_v :

$$C_p > C_v.$$

At first sight it might appear that this inequality is due simply to the work which must be done by a body in expanding on heating at constant pressure. This is not so, however; the inequality applies also to the few substances which contract on heating, as well as to those which expand. It is in fact a consequence of a very general theorem of thermodynamics: an external interaction which removes a body from a state of thermal equilibrium brings about processes in it which, as it were, try to reduce the effect of this interaction. For example, heating a body brings about processes which absorb heat, whereas cooling brings about processes in which heat is evolved. This is called *Le Chatelier's principle*.

Let us imagine that a body in equilibrium with an external medium receives a quantity of heat in such a way that its volume remains unchanged and its temperature increases by an amount $(\Delta T)_v$. The pressure of the body will also be changed, and the equilibrium condition, according to which this pressure must be equal to that of the surrounding medium, will no longer be satisfied. According to Le Chatelier's principle, the restoration of equilibrium, which would restore the original pressure, must be accompanied by cooling. In other words, the change $(\Delta T)_p$ in the temperature of the body at constant pressure is less than the change $(\Delta T)_v$ at constant volume (for a given quantity of heat gained by the body). This means that, for a given change in temperature, more heat is necessary at constant pressure than at constant volume.

In what follows we shall several times make use of Le Chatelier's principle to decide the direction in which a quantity changes when another quantity is varied.

§57. The specific heat of gases

Since the molecules of an ideal gas are assumed not to interact with one another, the change in their mean distance apart when the volume of the gas varies cannot affect its internal energy. In other words, the internal energy of an ideal gas is a function only of its temperature, and not of its volume or pressure. Hence the specific heat $C_v = dE/dT$ of the gas also depends only on the temperature.

The same is true of the specific heat $C_p = dW/dT$, and there is a very simple relation between the two specific heats of the gas. From the equation of state $pV = RT$, the enthalpy of one mole of gas is related to its internal energy by

$$W = E + pV = E + RT.$$

Differentiating this expression with respect to temperature, we obtain

$$C_p = C_v + R,$$

i.e. the difference of the molar specific heats of the gas, $C_p - C_v$, is equal to the gas constant $R = 8.3 \text{ J/deg.mole} = 2 \text{ cal/deg.mole}$.

It is easy to find the specific heat of a monatomic gas (such as the noble gases). In this case the internal energy of the gas is simply the sum of the kinetic energies of the translational motion of the particles. Since, by the definition of temperature, the mean kinetic energy of one particle is $\frac{3}{2}kT$, the internal energy of one mole of gas is

$$E = \frac{3}{2}N_0kT = \frac{3}{2}RT.$$

The specific heats are therefore

$$C_V = \frac{3}{2}R = 12.5 \text{ J/deg.mole},$$

$$C_p = \frac{5}{2}R = 20.8 \text{ J/deg.mole}.$$

These values are quite independent of temperature.

We shall see later that in many processes an important property of the gas is the ratio of the specific heats C_p and C_V , usually denoted by γ :

$$\gamma = C_p/C_V.$$

For monatomic gases

$$\gamma = 5/3 = 1.67.$$

The specific heat of diatomic and polyatomic gases is more complicated than that of monatomic gases. Their internal energy consists of the kinetic energies of translation and rotation of the molecules and the energy of the atoms vibrating within the molecule. Thus each of these three types of motion makes a certain contribution to the specific heat of the gas.

Here we may return to the definition of temperature given in §50. Since a molecule has three degrees of freedom in its translational motion, we can say that each of them corresponds to a mean kinetic energy $\frac{1}{2}kT$. According to classical mechanics, the same result would be obtained for every degree of freedom of the molecule, whether for translational motion, rotation, or vibration of the atoms within it. We know also that in the vibrational motion the mean value of the potential energy is equal to the mean value of the kinetic energy. Thus, according to classical

mechanics, the thermal potential energy of each degree of freedom of the vibration of atoms within the molecule would also be $\frac{1}{2}kT$. Thus we find that any gas should have a constant specific heat independent of temperature and determined entirely by the number of degrees of freedom of the molecule (and therefore by the number of atoms in it).

In reality, however, the vibrational motion of the atoms in the molecule affects the specific heat of the gas only at sufficiently high temperatures. The reason is that this motion remains of the nature of “zero-point vibrations”, not only at low temperatures but also at comparatively high temperatures, on account of the comparatively large energy of these vibrations. The “zero-point energy”, by its nature, is independent of temperature, and therefore does not affect the specific heat. For example, in the molecules of diatomic gases (nitrogen, oxygen, hydrogen etc.), the vibrations of the atoms within the molecules are fully “included” in the motion only at temperatures of the order of thousands of degrees; at lower temperatures their contribution to the specific heat decreases rapidly and is practically zero even at room temperature.

The zero-point energy of rotation of molecules is very small, and thus classical mechanics is very soon applicable to this motion: at temperatures of a few degrees Kelvin for diatomic molecules, with the exception of the lightest gas, hydrogen, for which a temperature of about 80°K is necessary.

In the neighbourhood of room temperature, the specific heat of diatomic gases is therefore due only to the translational and rotational motion of the molecules and is very close to its theoretical constant value (in classical mechanics)

$$C_v = \frac{5}{2}R = 20.8 \text{ J/deg.mole,}$$

$$C_p = \frac{7}{2}R = 29.1 \text{ J/deg.mole.}$$

The ratio of specific heats $\gamma = 7/5 = 1.4$.

We may note that in the “quantum region” the mean energies of the thermal rotational and vibrational motions, and therefore the specific heat of the gas, depend not only on the temperature but also on the “individual” properties (moments of inertia and vibrational frequencies) of the molecule. [It is for this reason

that these energies, unlike the energy of the translational motion, cannot be used for a direct definition of temperature.]

The specific heat of polyatomic gases is even more complicated. The atoms in a polyatomic molecule can execute oscillations of various types with various zero-point energies. As the temperature rises, these oscillations are successively "included" in the thermal motion, and the specific heat of the gas increases accordingly. It may happen, however, that the inclusion of all the oscillations is never achieved, since the molecules may disintegrate at high temperatures.

It may again be recalled that the whole of the above discussion is for the case of an ideal gas. At high pressures, when the properties of the gas become appreciably different from those of an ideal gas, its specific heat is also changed, because of the contribution to the internal energy arising from the interaction between the molecules.

§58. Solids and liquids

The simplicity of the thermal properties of an ideal gas, which allows a general equation of state for all gases to be derived, is due to the fact that the interaction between molecules in the gas is unimportant. In solids and liquids, the interaction between the molecules is of primary importance; the thermal properties of these substances therefore differ considerably, and it is impossible to establish any general equation of state.

Solids and liquids, unlike gases, are not readily compressed. The *compressibility* of a substance is usually defined as

$$\kappa = -\frac{1}{V} \left(\frac{dV}{dp} \right)_T;$$

the derivative of the volume with respect to the pressure is taken at constant temperature, i.e. describes a process of isothermal compression. This coefficient is negative, i.e. the volume decreases when the pressure increases, and the minus sign is used in order to make the compressibility a positive quantity. The dimensions of κ are evidently the reciprocal of those of pressure.

As examples, we may give the values of the compressibility per bar for various liquids at room temperature and atmospheric pressure:

Mercury	$0.4 \times 10^{-5} \text{ bar}^{-1}$	Alcohol	$7.6 \times 10^{-5} \text{ bar}^{-1}$
Water	4.9×10^{-5}	Ether	14.5×10^{-5}

The compressibilities of most solids are even smaller:

Diamond	$0.16 \times 10^{-6} \text{ bar}^{-1}$	Aluminium	$1.4 \times 10^{-6} \text{ bar}^{-1}$
Iron	0.61×10^{-6}	Glass	2.7×10^{-6}
Copper	0.76×10^{-6}	Caesium	62×10^{-6}

For comparison, let us find the compressibility of a gas. In isothermal compression, the volume of a gas decreases in inverse proportion to the pressure: $V = RT/p$. Substituting this expression in the foregoing definition of the compressibility κ , we have after the differentiation

$$\kappa = 1/p.$$

At a pressure of 1 bar, the compressibility of the gas is 1 bar^{-1} .

Another quantity used to describe the thermal properties of solids and liquids is the *coefficient of thermal expansion* defined as

$$\alpha = \frac{1}{V} \left(\frac{dV}{dT} \right)_p;$$

the suffix p to the derivative means that the body is heated at constant pressure.

The majority of bodies expand on heating, and the coefficient α is positive. This is to be expected, since the greater thermal motion tends to move the molecules apart. Nevertheless, there are exceptions to this rule. For example, water contracts on heating in the range from 0 to 4°C . Liquid helium also contracts on heating at temperatures below 2.19°K (helium II; see §74).

As examples, we may give the coefficients of thermal expansion of various liquids at room temperature:

Mercury	$1.8 \times 10^{-4} \text{ deg}^{-1}$	Alcohol	$10.8 \times 10^{-4} \text{ deg}^{-1}$
Water	2.1×10^{-4}	Ether	16.3×10^{-4}

[For comparison, the coefficient of thermal expansion of gases,

obtained by substituting $V = RT/p$ in the definition of α , is $\alpha = 1/T$; for $T = 293^\circ\text{K}$, $\alpha = 3.4 \times 10^{-3} \text{ deg}^{-1}$.]

The coefficient of thermal expansion of solids is still smaller:

Iron	$3.5 \times 10^{-5} \text{ deg}^{-1}$
Copper	5.0×10^{-5}
Glass	2.4 to 3.0×10^{-5}

Invar (an alloy of 64% iron and 36% nickel) and fused quartz have especially small values of α (3×10^{-6} and 1.2×10^{-6} respectively). These substances are widely used in making parts of instruments in which it is desirable to avoid dimensional changes when the temperature varies.

It has been mentioned in §45 that the thermal expansion of crystals (other than cubic) occurs differently in different directions. This difference may be very considerable. For example, in the thermal expansion of a crystal of zinc, the linear dimension in the direction of the hexagonal axis increases 4.5 times faster than those in the directions perpendicular to this axis.

The specific heat of solids and liquids, like that of gases, usually increases with temperature. The specific heat of a solid depends on the energy of atoms executing small thermal oscillations about their equilibrium positions. When the temperature rises, this specific heat tends to a certain limit corresponding to the state where the oscillations of the atoms can be treated on the basis of classical mechanics. Since the motion of the atoms is entirely oscillatory, a mean energy kT must correspond to each of its three degrees of freedom: $\frac{1}{2}kT$ from the mean kinetic energy and $\frac{1}{2}kT$ from the mean potential energy (as described in §57). The total mean energy per atom in a solid would then be $3kT$.

This limit, however, is never reached for compounds of any complexity, since the substance melts or decomposes before this occurs. At ordinary temperatures the limiting value of the specific heat is reached for many elements, so that the specific heat of one gram-atom of a solid element is approximately

$$C = 3R = 25 \text{ J/deg.mole} = 6 \text{ cal/deg.mole};$$

this is sometimes called *Dulong and Petit's law*.

In discussing the specific heat of a solid we deliberately do not distinguish between the specific heats at constant pressure and

at constant volume. The measured specific heats are usually those at constant pressure, but in solids the difference between C_p and C_V is very small (e.g. for iron $C_p/C_V = 1.02$). This is because of the smallness of the coefficient of thermal expansion for solids: there is a general relation between the difference of specific heats for any body, the coefficient of thermal expansion α and the compressibility κ :

$$c_p - c_V = T\alpha^2/\rho\kappa,$$

where ρ is the density of the substance and c_p and c_V the specific heats per gram. Thus we see that the difference $c_p - c_V$ is proportional to the square of the coefficient α .

As the temperature decreases, the specific heat of a solid also decreases and tends to zero at absolute zero. This is a consequence of a remarkable general theorem (called *Nernst's theorem*), according to which, at sufficiently low temperatures, any quantity representing a property of a solid or liquid becomes independent of temperature. In particular, as absolute zero is approached, the energy and enthalpy of a body no longer depend on the temperature; the specific heats c_p and c_V , which are the derivatives of these quantities with respect to temperature, therefore tend to zero.

It also follows from Nernst's theorem that, as $T \rightarrow 0$, the coefficient of thermal expansion tends to zero, since the volume of the body ceases to depend on the temperature.

THERMAL PROCESSES

§59. Adiabatic processes

Let us now consider some simple thermal processes. A very simple process is the expansion of a gas into a vacuum: the gas is initially in a part of a vessel separated from the rest of the vessel by a partition, and then an opening is made in the partition and the gas fills the whole vessel. Since the gas does no work in such an expansion, its energy remains constant, i.e. the energy E_1 of the gas before the expansion is equal to its energy E_2 after the expansion:

$$E_1 = E_2.$$

For an ideal gas the energy depends, as we know, only on the temperature; thus, since the energy is constant, it follows that the temperature of an ideal gas remains constant when it expands into a vacuum. However, the temperature of gases which are not nearly ideal changes on expansion into a vacuum.

There is another process of expansion of a gas called an *adiabatic* process, which differs very greatly from expansion into a vacuum. Adiabatic processes are of great importance, and will now be considered in detail.

The typical feature of an adiabatic process is that the gas remains continuously under an external pressure equal to the pressure of the gas itself. Another condition for an adiabatic process is that throughout the process the gas remains thermally isolated from the external medium, i.e. does not gain or lose heat.

It is simplest to imagine the adiabatic expansion (or compression) of a gas in a thermally isolated cylindrical vessel with a piston. When the piston is moved out sufficiently slowly, the gas expands behind it and at every instant has a pressure correspond-

ing to the total volume which it then occupies. Here “sufficiently slowly” means, therefore, so slowly that the gas is able to establish thermal equilibrium corresponding to every instantaneous position of the piston. If, on the other hand, the piston is moved out too rapidly, the gas will not be able to follow it, and a region of reduced pressure will exist beneath the piston, into which the remaining gas will expand; similarly, if the piston is moved in too rapidly, a region of increased pressure will exist. Such processes would not be adiabatic.

In practice, this condition of slowness is very easily fulfilled in the case considered. Analysis shows that the condition would not be fulfilled only if the rate of movement of the piston were comparable with the velocity of sound in the gas. Thus, in the practical carrying out of an adiabatic expansion, the principal condition is that of thermal isolation, which requires that the process should be “sufficiently fast”: the gas must not be able to exchange heat with the external medium during the process. It is clear that this condition is entirely compatible with the condition of “sufficient slowness” stated above; it depends on the thoroughness of the thermal isolation of the vessel and may be said to be of secondary importance and unrelated to the actual nature of the process. For this reason an adiabatic process is regarded in physics as one which primarily satisfies the condition of “sufficient slowness”, the latter being fundamental. We shall return to a discussion of this condition in §62.

In an adiabatic process we can not say that the internal energy of the gas itself remains constant, since the gas does work when it expands (or work is done on it when it is compressed). The general equation of an adiabatic process is obtained by putting the quantity of heat dQ equal to zero in the relation $dQ = dE + pdV$, in accordance with the condition of thermal isolation. Thus an infinitesimal change in the state of a body in an adiabatic process is described by the equation

$$dE + pdV = 0.$$

Let us apply this equation to the adiabatic expansion (or compression) of an ideal gas; for simplicity, all quantities will refer to one mole. The energy of an ideal gas is a function only of its temperature, and the derivative dE/dT is the specific heat

C_V ; in the equation of the adiabatic process, we can therefore replace dE by $C_V dT$:

$$C_V dT + p dV = 0.$$

Substituting $p = RT/V$ and dividing the equation by T , we obtain the relation

$$C_V dT/T + R dV/V = 0.$$

Let us assume further that the specific heat of the gas is constant in the temperature range considered; for monatomic gases this is always true, and for diatomic gases it is true over a wide range of temperatures. Then the above relation may be written

$$d(C_V \log_e T + R \log_e V) = 0,$$

whence

$$C_V \log_e T + R \log_e V = \text{constant}$$

or, in power form,

$$T^{C_V} V^R = \text{constant}.$$

Finally, since for an ideal gas $C_p - C_V = R$, the $1/C_V$ power of this equation may be written

$$TV^{\gamma-1} = \text{constant},$$

where $\gamma = C_p/C_V$.

We see that in an adiabatic process the temperature and volume of an ideal gas vary in such a way that the product $TV^{\gamma-1}$ remains constant. Since γ is always greater than unity, $\gamma - 1 > 0$, and therefore an adiabatic expansion is accompanied by a cooling of the gas, and an adiabatic compression by heating.

On combining the above equation with the formula $pV = RT$ we can derive a similar relation between the temperature and the pressure in an adiabatic process:

$$Tp^{-(\gamma-1)/\gamma} = \text{constant},$$

and the relation

$$pV^\gamma = \text{constant}$$

between the pressure and the volume; this last relation is called the equation of *Poisson's adiabatic*.

In isothermal expansion of a gas, its pressure decreases in inverse proportion to the volume V . In adiabatic expansion, we see that the pressure decreases in inverse proportion to V^γ , i.e. more rapidly (since $\gamma > 1$). If these processes are represented graphically by plotting p against V as two curves, an isothermal and an adiabatic, intersecting at a point p_0, V_0 which represents the initial state of the gas, then the adiabatic curve will be steeper than the isothermal (Fig. 78).

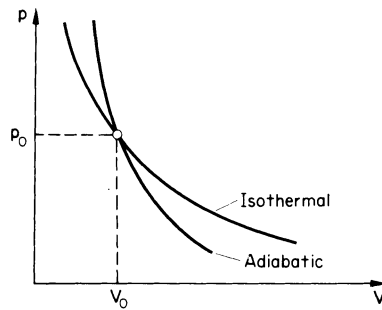


FIG. 78.

This property may be stated in another manner by considering the change in volume as a function of pressure (i.e. by turning Fig. 78 through 90°) and representing this relation by the compressibility $\kappa = -(1/V)dV/dp$; see §58, where the isothermal case was considered. It is then easy to see that the adiabatic compressibility of a gas is less than its isothermal compressibility:

$$\kappa_{\text{ad}} < \kappa_{\text{is}}.$$

This inequality, derived here for gases, is in fact valid for all bodies, and follows from Le Chatelier's principle.

On the other hand, another property of adiabatic processes in a gas, the heating on compression, is not a universal property of adiabatic compression of all bodies. This is likewise seen from Le Chatelier's principle. If a body is compressed without gaining any heat (which itself would affect the temperature of the body), the temperature of the body will change so as to oppose the compression. For the great majority of bodies, which expand on heating, this means that the temperature will rise on adiabatic compression (and conversely will fall on expansion). But it is clear from this discussion that, if the volume of a body decreases on heating, an adiabatic compression of the body will be accompanied by cooling.

§60. Joule-Kelvin processes

Processes in which a gas or liquid passes steadily from one pressure to another without exchange of heat with the surrounding medium are of considerable interest. By "steadily" we here mean that the two pressures remain constant throughout the process.

Such a process is in general accompanied by a flow of gas (or liquid) with some velocity different from zero, but this velocity can be made very small by causing the gas to go from one pressure to the other through an obstruction which greatly impedes the flow, such as a porous partition or a small hole.

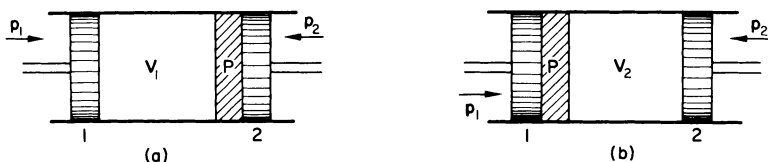


FIG. 79.

The steady passage of a thermally isolated gas from one pressure to another under conditions where the gas does not acquire any appreciable velocity is called a *Joule-Kelvin process*. This process may be diagrammatically represented by the passage of a gas in a cylindrical vessel through a porous partition P (Fig. 79a, b), the pressures p_1 and p_2 on each side of the partition being maintained constant by pistons 1 and 2.

Let the gas initially occupy a volume V_1 between piston 1 and the partition P (Fig. 79a). Piston 1 is now moved in and piston 2 moved out, keeping the pressures p_1 and p_2 acting on the pistons unchanged. The gas, passing at a low velocity through the porous partition, will finally occupy a volume V_2 between the partition and piston 2, and will be at a pressure p_2 (Fig. 79b).

Since in this process there is no exchange of heat with the surrounding medium, the work done by the pistons must be equal to the change in the internal energy of the gas. The gas pressures remain constant during the process, and therefore the work done by the piston 1 in displacing the gas from the volume V_1 is simply the product p_1V_1 . The gas passing through the partition does work on the piston 2. Thus the total work done by the pistons on the gas is $p_1V_1 - p_2V_2$, and this, as already stated, must be equal to the increase in the internal energy of the gas:

$$p_1V_1 - p_2V_2 = E_2 - E_1,$$

where E_1 and E_2 are the internal energies of a given quantity of the gas in the initial and final states. Hence

$$E_1 + p_1V_1 = E_2 + p_2V_2,$$

or

$$W_1 = W_2,$$

where $W = E + pV$ is the enthalpy. Thus the enthalpy of the gas is conserved in a Joule-Kelvin process.

For an ideal gas both the energy and the enthalpy depend only on the temperature. Thus the equality of the enthalpies implies that the temperatures are equal: if an ideal gas undergoes a Joule-Kelvin process, its temperature remains unchanged.

In real gases the temperature changes in a Joule-Kelvin process, and may do so by a considerable amount. For example, when air at room temperature expands from 200 atm pressure to 1 atm, it is cooled by about 40°.

At sufficiently high temperatures, all real gases are heated by expansion in a Joule-Kelvin process, while at lower temperatures (and not too high pressures) they are cooled; there is therefore a

temperature (called the *inversion point*) above which the change in temperature in a Joule–Kelvin process is of opposite sign. The position of the inversion point depends on the pressure, and is different for different gases. For example, air is cooled in a Joule–Kelvin process at room temperature, but to achieve this effect in hydrogen it must first be cooled to about 200°K or below, and for helium a temperature of 40°K is necessary.

The change in temperature in a Joule–Kelvin process is widely used in technology for the liquefaction of gases. The gas velocity is usually lowered by means of a narrow opening called an expansion valve.

§61. Steady flow

In a Joule–Kelvin process, the gas passes steadily from one pressure to the other, and its velocity is artificially made small by means of friction. However, the results obtained by considering this process are easily generalised to the case of any steady thermally isolated flow of gas (or liquid) with non-zero velocity. The only difference is that the kinetic energy of the flowing gas can not now be neglected. The work done on the gas increases its energy, which now includes the kinetic energy of its motion as a whole as well as its internal energy. Thus, for a steady flow of gas or liquid we have

$$\frac{1}{2}Mv^2 + E + pV = \text{constant}$$

or

$$\frac{1}{2}Mv^2 + W = \text{constant},$$

where W and M are the enthalpy and the mass of a given quantity of substance and v the velocity of flow. The above equation signifies that the quantity $\frac{1}{2}Mv^2 + W$ is the same for a given mass of substance no matter where it is in the flow.

Where it may be necessary to take into account also the potential energy in a field of gravity in the flow of a liquid (the weight is unimportant in gas flow), we can similarly write

$$\frac{1}{2}Mv^2 + Mgz + E + pV = \text{constant},$$

where z is the height of a given point in the flow.

Let us assume that the motion in the flow is not accompanied by any appreciable friction, either within the flowing substance

itself or against any external obstacles; this is in a sense the opposite of a Joule–Kelvin process, where friction plays an important part. Under these conditions we can assume not only that the flow as a whole is thermally isolated from the external medium (as we have assumed throughout) but also that during the motion each individual element of substance is thermally isolated; if there were appreciable friction this would not be so, since frictional heat would be generated within the flow. That is, we may assume that during the motion each element of substance expands or contracts adiabatically.

Let us consider, for example, the outflow of gas under these conditions from a vessel in which it is at a pressure p different from the atmospheric pressure p_0 . If the outflow takes place through a sufficiently small opening, the velocity of the gas within the vessel may be taken as zero. The velocity v of the outflowing jet is given by the equation

$$W_0 + \frac{1}{2}v^2 = W;$$

here we have taken the mass M as 1 g, so that W and W_0 are the enthalpies per gram of gas within the vessel and in the outflowing jet. If the gas is assumed ideal and its specific heat independent of temperature, then the formula $c_p = dW/dT$ or $dW = c_p dT$ (cf. § 56) shows that $W_0 - W = c_p(T_0 - T)$, and hence

$$v^2 = 2c_p(T - T_0).$$

Finally, the temperature T_0 in the outflowing jet can be expressed in terms of the temperature T of the gas in the vessel by means of the equation of adiabatic expansion of the gas derived in §59; this states that the product $Tp^{-(\gamma-1)/\gamma}$ is constant:

$$T_0 = T(p_0/p)^{(\gamma-1)/\gamma}.$$

Thus we finally obtain the following formula for the velocity of outflow of the gas:

$$v^2 = 2c_p T [(p_0/p)^{(\gamma-1)/\gamma} - 1].$$

The flow of liquids generally occurs without any appreciable change in their volume, owing to their comparatively small

compressibility. In other words, a flowing liquid may be regarded as incompressible and of constant density.

The equation of steady (frictionless) flow of such a liquid is especially simple. In this case the general equation of an adiabatic process, $dE + pdV = 0$, reduces to $dE = 0$ simply, since $dV = 0$ owing to the incompressibility of the liquid. That is, the energy E remains constant and may therefore be omitted from the left-hand side of the equation

$$\frac{1}{2}Mv^2 + E + pV + Mgz = \text{constant.}$$

Dividing this equation by the mass M and noting that the ratio M/V is the density ρ of the liquid, we finally deduce that the following quantity remains constant throughout a thermally isolated steady frictionless flow of an incompressible liquid:

$$\frac{1}{2}v^2 + p/\rho + gz = \text{constant.}$$

This is called *Bernoulli's equation*.

As an example, let us consider the motion of a liquid in a pipe of variable cross-section, which for simplicity we shall assume to lie horizontally. Then the force of gravity has no effect on the motion, and Bernoulli's equation gives

$$\frac{1}{2}v^2 + p/\rho = \frac{1}{2}v_0^2 + p_0/\rho,$$

where v_0 and v are the flow velocities at any two cross-sections of the tube, and p_0 and p the corresponding pressures. If the areas of these two cross-sections are S_0 and S , the volumes of liquid passing through them per unit time are v_0S_0 and vS , and since the liquid is assumed incompressible $vS = v_0S_0$, or

$$v = v_0S_0/S,$$

i.e. the velocity of an incompressible liquid at any cross-section is inversely proportional to its area. Substituting this expression for v in Bernoulli's equation, we obtain a relation between the pressure and the cross-sectional area:

$$\begin{aligned} p &= p_0 + \frac{1}{2}\rho(v_0^2 - v^2) \\ &= p_0 + \frac{1}{2}\rho v_0^2(1 - S_0^2/S^2). \end{aligned}$$

We see that the pressure is greater in the wider parts of the pipe than in the narrower ones.

Let us now apply Bernoulli's equation to determine the velocity of a jet of liquid leaving a vessel through a small opening. Since the area of the opening is assumed small in comparison with the cross-section of the vessel, we may neglect the fall in the level of the liquid in the vessel. Using also the fact that the pressure on the surface of the liquid in the vessel and the pressure in the jet are the same, and equal to the atmospheric pressure, we obtain from Bernoulli's equation

$$\frac{1}{2}v^2 + gz_1 = gz_2,$$

where v is the velocity of the outflowing jet, and z_2 and z_1 the heights of the surface of the liquid in the vessel and the point of outflow of the liquid; hence

$$v = \sqrt{2gh},$$

where $h = z_2 - z_1$. This formula, called *Torricelli's formula*, shows that the velocity of the outflowing liquid from a small aperture is the same as the velocity of fall of a body from a height h which is equal to the height of the liquid in the vessel above the aperture.

§62. Irreversibility of thermal processes

The mechanical movements of material bodies, occurring in accordance with the laws of mechanics, have the following remarkable property. Whatever the motion of a body, the reverse motion is always possible, i.e. the motion in which the body passes through the same points in space with the same velocities as in the original motion, but in the opposite direction. For example, let a body be projected in the field of gravity at a certain angle to the horizontal; it will describe a certain trajectory and fall to the ground at some point. If now the body is projected from this point at the angle at which it fell and at the corresponding velocity, it will describe the same trajectory in the opposite direction and fall to its original position (if air friction may be neglected).

This *reversibility* of mechanical motions may be alternatively formulated by saying that they are symmetrical as regards inter-

changing the future and the past, i.e. with respect to time reversal. The symmetry of mechanical motions follows at once from the equations of motion themselves, since when the sign of the time is reversed so is that of the velocity, but the acceleration is left unchanged.

The situation is quite different as regards thermal phenomena. If a thermal process takes place, then the reverse process (i.e. the process in which the same thermal states are traversed in the opposite order) is in general impossible. Thus thermal processes are as a rule *irreversible*.

For example, if two bodies at different temperatures are brought into contact, the hotter body will transmit heat to the colder body, but the reverse process (a spontaneous direct transfer of heat from the colder to the hotter body) never occurs.

The expansion of a gas into a vacuum, described in §59, is likewise an irreversible process. The gas spreads through the opening on both sides of the partition, but without external interference it will never collect spontaneously in one half of the vessel again.

Any system of bodies left to itself tends to reach a state of thermal equilibrium, in which the bodies are at relative rest, with equal temperatures and pressures. Having reached such a state, the system will not of its own accord leave that state. In other words, all thermal phenomena accompanied by processes of approach to thermal equilibrium are irreversible.

For instance, all processes accompanied by friction between moving bodies are irreversible. The friction causes a gradual slowing down of the motion (the kinetic energy being converted into heat), i.e. an approach to a state of equilibrium in which there is no motion. For this reason, in particular, a Joule–Kelvin process, in which the gas passes through an obstacle with a large amount of friction, is irreversible.

All thermal processes occurring in Nature are to some extent irreversible. In some cases, however, the degree of irreversibility may be so slight that the process may be regarded as reversible with sufficient accuracy.

It is clear from the foregoing that, in order to achieve reversibility, it is necessary to eliminate from the system as far as possible all processes which constitute an approach to thermal equilibrium. For example, there must be no direct transfer of

heat from a hotter to a colder body and no friction in the motion of bodies.

An example of a process which is reversible to a high degree (and, in the ideal case, perfectly reversible) is the adiabatic expansion or compression of a gas described in §59. The condition of thermal isolation excludes a direct exchange of heat with the surrounding medium. The “sufficiently slow” movement of the piston ensures that there are no irreversible processes of expansion of a gas into the vacuum which would be produced behind a too rapidly moving piston, since this is what the condition of slowness signifies. Of course, in practice there will still remain some causes of irreversibility (imperfect thermal isolation of the vessel containing the gas; friction in the movement of the piston).

“Slowness” is a general characteristic of reversible processes: the process must be so slow that the bodies involved in it are able to reach at every instant the state of equilibrium which corresponds to the prevailing external conditions. In the example of the expansion of a gas, the latter must be able to follow the piston and remain homogeneous throughout its volume. Complete reversibility could be achieved only in the ideal case of an infinitely slow process, and for this reason alone a process occurring at a finite rate cannot be completely reversible.

We have already mentioned that, in a system of bodies in thermal equilibrium, no process can take place without external interference. This can be stated in another way: bodies in thermal equilibrium can do no work, since work requires mechanical motion, i.e. a conversion of energy into the kinetic energy of the bodies.

This extremely important assertion that work can not be obtained from the energy of bodies in thermal equilibrium is called the *second law of thermodynamics*. We are always surrounded by considerable sources of thermal energy in a state close to equilibrium. An engine working merely on the energy of bodies in thermal equilibrium would constitute a perpetual-motion machine. The second law of thermodynamics prevents the construction of such a *perpetual-motion machine of the second kind*, just as the *first law of thermodynamics* (the law of conservation of energy) prevents that of a machine of the first kind, i.e. one which would do work “from nothing”, without any external source of energy.

§63. The Carnot cycle

From the foregoing it follows that work can be done only by means of a system of bodies which are not in thermal equilibrium with one another. Let us imagine such a system idealised as two bodies at different temperatures. If the two bodies are simply brought into contact, heat will pass from the hotter to the colder body, but no work will be done. The transfer of heat from a hotter to a colder body is an irreversible process, and this example demonstrates the general rule that irreversible processes prevent the doing of work.

If it is desired to obtain the maximum possible work from given bodies, the process must be made as nearly reversible as possible: all irreversible processes must be avoided, and only processes which occur to the same extent in both directions must be used.

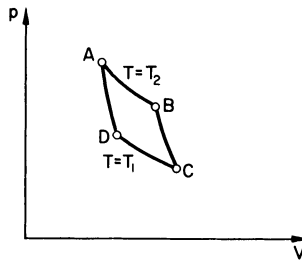


FIG. 80.

Returning to the system of the two bodies, we denote their temperatures by T_1 and T_2 (and let $T_2 > T_1$), and conventionally call the hotter body a heat source, and the colder body a heat sink. Since direct exchange of heat between these bodies is not permissible, it is clear first of all that, in order to do work, a further body must be used; this will be called the working medium. It may be imagined as a cylindrical vessel containing gas and closed by a piston.

We shall represent the process in which the medium takes part, using a pV diagram (Fig. 80). Let the gas be initially at a temperature T_2 , and let its state be represented by the point A in the diagram. The working medium is now brought into contact with the heat source, and the gas is caused to expand; it gains a

certain quantity of heat from the heat source, while remaining at the source temperature T_2 (the total quantity of heat in the heat source is assumed so large that its temperature is not changed when a small quantity of heat is transferred to the gas). Thus the gas undergoes a reversible isothermal expansion, since heat is transferred only between bodies at the same temperature. In Fig. 80 this process is shown by the isotherm AB .

Next, the working medium is removed from the heater, thermally isolated and further expanded, this time adiabatically. In this expansion the gas is cooled, and the expansion is continued until the temperature of the gas falls to the temperature T_1 of the heat sink. This process is represented in the diagram by the adiabatic BC , which is steeper than the isotherm AB , since in adiabatic expansion the pressure falls more rapidly than in isothermal expansion.

The working medium is now brought into contact with the heat sink and the gas is isothermally compressed at temperature T_1 , thereby transferring a certain quantity of heat to the heat sink. Finally, the working medium is removed from the heat sink and the gas is adiabatically compressed to return it to its initial state; for this purpose it is necessary to make the proper choice of the point D , i.e. the volume to which the isothermal compression CD is taken.

Thus the working medium undergoes a cyclic process, returning to its original state but doing a certain quantity of work represented by the area of the curvilinear quadrilateral $ABCD$. This work is done by virtue of the fact that on the upper isotherm the working medium takes from the heat source a greater quantity of heat than it gives to the heat sink on the lower isotherm. Every stage of this cyclic process is reversible, and the work done is therefore the maximum possible for a given quantity of heat taken from the source.

The process just described is called a *Carnot cycle*. It shows that, in principle, work can be done reversibly by means of two bodies at different temperatures. Being the maximum possible amount, this work is independent of the properties of the working medium.

The ratio of the work done to the quantity of energy taken from the hotter body is called the *efficiency of the heat engine* and will be denoted by η . It is clear from the above that the

efficiency of a Carnot cycle is the maximum possible for any heat engine operating with given temperatures of the heat source and sink. It can be shown (see §65) that this efficiency is

$$\eta_{\max} = (T_2 - T_1)/T_2.$$

Thus, even in the ideal limit of completely reversible operation of a heat engine, the efficiency is less than unity: a fraction T_1/T_2 of the energy taken from the heat source is transferred unprofitably to the heat sink as heat. This fraction decreases with increasing temperature T_2 for given T_1 . The temperature T_1 is usually that of the surrounding air, and therefore cannot be reduced. To decrease the fraction of energy wasted, therefore, the aim in applications is to operate the engine at the maximum possible temperature T_2 .

The efficiency of an actual heat engine is always less than η_{\max} because of the irreversible processes which unavoidably occur in it. The quantity η/η_{\max} , i.e. the ratio of the efficiency of the actual engine to that of an ideal engine with the same heat source and sink temperatures, can be used to represent the degree to which the engine approaches the ideal one. This is therefore the ratio of the work done by the heat engine to the maximum work which could be obtained in the given conditions if the engine were operating reversibly.

§64. The nature of irreversibility

All thermal phenomena reduce ultimately to the mechanical movement of the atoms and molecules in a body. The irreversibility of thermal processes is therefore, at first sight, in conflict with the reversibility of all mechanical motions. This contradiction is in fact only apparent.

Suppose that a body slides on another body. Because of friction, this motion will be gradually slowed down and the system will finally reach a state of thermal equilibrium; the motion will then cease. The kinetic energy of the moving body is converted into heat in this process, i.e. kinetic energy of the random motion of the molecules in both bodies. This conversion of energy into heat can obviously be brought about in an infinite number of ways: the kinetic energy of the motion of the body as a whole can be distributed between the enormous number of

molecules in an enormous number of ways. In other words, the state of equilibrium in which there is no macroscopic motion can occur in an immensely greater number of ways than a state in which a considerable quantity of energy is concentrated in the form of kinetic energy of the ordered motion of the body as a whole.

Thus the change from a non-equilibrium state to an equilibrium state is a change from a state which can occur in a small number of ways to one which can occur in a very much larger number of ways. It is clear that the most probable state of a body (or system of bodies) is that which can occur in the largest number of ways, and this will be the state of thermal equilibrium. Thus, if a system left to itself (i.e. a closed system) is not in a state of equilibrium, then its subsequent behaviour will almost certainly be to enter a state which can occur in a very large number of ways, i.e. to approach equilibrium.

On the other hand, when a closed system has reached a state of equilibrium, it is most unlikely to leave that state spontaneously.

Thus the irreversibility of thermal processes is *probabilistic*. The spontaneous passage of a body from an equilibrium state to a non-equilibrium state is, strictly speaking, not impossible, but only very much less probable than that from a non-equilibrium state to an equilibrium state. The irreversibility of thermal processes is ultimately due to the very large number of molecules of which bodies are composed.

The improbability of a body's spontaneously leaving an equilibrium state may be judged by considering the expansion of a gas into a vacuum. Let the gas be initially in one half of a vessel divided by a partition into two equal parts. When an opening is made in the partition, the gas spreads uniformly through both parts of the vessel. The opposite transfer of the gas into one half of the vessel will never occur without external interference. The reason for this is easily seen by a simple calculation. Each molecule of gas, in its motion, spends on average the same time in each part of the vessel; we may say that the probability of finding it in either half of the vessel is $\frac{1}{2}$. If the gas may be regarded as ideal, its molecules move independently. The probability of finding two given molecules in the same half of the vessel at the same time is therefore $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$; the probability of finding all N molecules of gas in one half of the vessel is 2^{-N} . For instance,

with a relatively small quantity of gas, containing say 10^{20} molecules, this probability is given by the fantastically small number $2^{-10^{20}} \approx 10^{-3 \times 10^{19}}$. In other words, this occurrence would be observed about once in a time represented by the number $10^{3 \times 10^{19}}$ —whether seconds or years is immaterial, since a second, a year, and indeed the time the Earth has existed, are equally small in comparison with this vast interval of time.

A similarly small number ($10^{-3 \times 10^{19}}$) may be shown to represent the probability that a single erg of heat will pass from a body at 0°C to another body at 1°C .

It is clear from these examples that the possibility of any appreciable spontaneous reversal of a thermal process is in essence a pure abstraction: its probability is so small that the irreversibility of thermal processes may in practice be regarded as exactly true.

The probabilistic nature of irreversibility appears, however, in the fact that in Nature there are nevertheless spontaneous deviations from equilibrium, although these are very small and short-lived; they are called *fluctuations*. Owing to fluctuations, for example, the density and temperature in different small regions of a body in equilibrium are not exactly constant, but undergo some very slight variations. For instance, the temperature of 1 milligram of water in equilibrium at room temperature will vary by amounts of the order of 10^{-8} degree. There are also phenomena in which fluctuations play an important part.

§65. Entropy

A quantitative characteristic of the thermal state of a body, which describes the degree to which it tends to enter other states, is the number of microscopic ways in which the state can occur. This number is called the *statistical weight* of the state and will be denoted by Γ . A body left to itself will tend to enter a state of greater statistical weight.

It is customary, however, to use instead of the number Γ itself its logarithm multiplied by Boltzmann's constant k . The quantity thus defined,

$$S = k \log_e \Gamma,$$

is called the *entropy* of the body.

The number Γ of ways in which a state of a system consisting of, for example, two bodies can occur is evidently equal to the product of the numbers Γ_1 and Γ_2 of ways in which the state of each body separately can occur: $\Gamma = \Gamma_1\Gamma_2$. Hence

$$\begin{aligned} S &= k \log_e \Gamma \\ &= k \log_e \Gamma_1 + k \log_e \Gamma_2 \\ &= S_1 + S_2. \end{aligned}$$

Thus the entropy of a composite system is equal to the sum of the entropies of its parts, and it is for this reason that the logarithm is used in the definition of the entropy.

The law which governs the direction of thermal processes may be formulated as a *law of increase of entropy*: in all thermal processes occurring in a closed system, the entropy of the system increases, and the maximum possible value of the entropy of a closed system is reached in a state of thermal equilibrium. This is a more precise quantitative form of the second law of thermodynamics. The law was stated by Clausius, and its interpretation in terms of molecular kinetics was given by Boltzmann.

Conversely, we may say that any process in which the entropy of a closed system increases is irreversible; the greater the increase in entropy, the higher the degree of irreversibility. The ideal case of a completely reversible process corresponds to that in which the entropy of a closed system remains constant.

A precise definition of what is meant by the "number of microscopic ways" in which a thermal state of a body can occur is given in *statistical physics*, and only when this has been done is it possible to carry out an actual calculation of the entropy of various bodies and to establish the relation between it and other thermal quantities.

A more detailed theoretical analysis makes it possible to derive a relation which is fundamental in thermodynamic applications of the concept of entropy. This relation is one between the change dS in the entropy of a body in an infinitesimal reversible change in state and the quantity of heat dQ which it gains in the process; the body is, of course, assumed not closed,

so that the reversibility of the process does not require that its entropy should be constant. The relation is

$$dS = dQ/T,$$

where T is the temperature of the body.

The existence of a relation between dS and dQ is entirely reasonable. When the body gains heat, the thermal motion of its atoms is increased, i.e. their distribution over various states of microscopic motion becomes more random, and so the statistical weight increases. It is also reasonable that the effect of a given quantity of heat on the thermal state of the body is described by the relative magnitude of this quantity of heat and the total internal energy of the body, and hence decreases with increasing temperature.

The relation $dQ = T dS$ leads, in particular, to the expression already given in §63 for the efficiency of a Carnot cycle. We have seen that this process involves three bodies: a heat source, a heat sink and a working medium. The latter is returned to its initial state by the cycle, and its entropy therefore also returns to its original value. The condition for the process to be reversible, i.e. the requirement that the total entropy of the system should be unchanged, therefore demands that the sum of the entropies S_1 of the heat sink and S_2 of the heat source should be constant. Let the sink gain a small quantity of heat ΔQ_1 in the cycle, and let the source lose ΔQ_2 . Then

$$\Delta S_1 + \Delta S_2 = \Delta Q_1/T_1 - \Delta Q_2/T_2 = 0,$$

whence $\Delta Q_1 = T_1 \Delta Q_2 / T_2$. The work done in one cycle is $A = \Delta Q_2 - \Delta Q_1$, and the efficiency is therefore

$$\eta = A/\Delta Q_2 = 1 - T_1/T_2.$$

PHASE TRANSITIONS

§66. Phases of matter

The evaporation of a liquid and the melting of a solid are processes of the type which are called in physics *phase transitions*. The characteristic feature of these processes is that they are discontinuous. For example, when ice is heated, its thermal state changes gradually until the temperature 0°C is reached and the ice suddenly begins to change into liquid water, which has entirely different properties.

States of matter between which phase transitions occur are called *phases* of matter. In this sense the aggregate states of matter (gaseous, liquid and solid) are different phases. For example, ice, liquid water and steam are the phases of water. The concept of phases, however, is broader than that of aggregate states; we shall see that different phases can exist within a single aggregate state.

It must be emphasised that, in speaking of the solid state as a separate phase of matter (distinct from the liquid phase), we are considering only the crystalline solid state. An amorphous solid is transformed on heating into a liquid by a gradual softening without discontinuity, as already described in §52; the amorphous solid state is therefore not a separate phase of matter. For instance, solid and liquid glass are not distinct phases.

The transition from one phase to another always occurs at a fixed temperature (at a given pressure). For example, ice begins to melt at 0°C (at atmospheric pressure) and on further heating the temperature remains constant until all the ice is changed into water. During this process, ice and water coexist in contact.

This exhibits another aspect of the temperature of a phase transition: it is the temperature at which there is thermal equilibrium between the two phases. In the absence of external interactions (including an external heat supply) the two phases

can coexist indefinitely at this temperature. At temperatures above or below the transition point, however, only one or the other phase can exist. For instance, at a temperature below 0°C only ice can exist (at atmospheric pressure), and above 0°C only liquid water.

When the pressure changes, so does the phase-transition temperature. In other words, a phase transition occurs when the pressure and temperature of the substance satisfy a certain fixed relation. This relation may be represented graphically as a curve in what is called a *phase diagram*, whose coordinates are the pressure p and the temperature T .

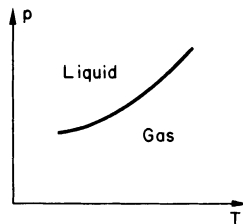


FIG. 81.

Let us consider, as an example, a phase transition between a liquid and its vapour. The phase-transition curve (called in this case the *evaporation curve*) determines the conditions under which the liquid and the vapour can coexist in equilibrium. The curve divides the plane into two parts, one of which corresponds to states of one phase and the other to states of the other phase (Fig. 81). Since in this case, at a given pressure, the higher temperatures correspond to the vapour and the lower ones to the liquid, the region to the right of the curve corresponds to the gaseous phase, and the region to the left corresponds to the liquid phase. The points on the curve itself correspond, as already mentioned, to states in which two phases coexist.

The phase diagram can be drawn not only in the pT plane but also in other coordinates: p and V , or T and V , where V is the volume of a given quantity of matter. We shall take V to be the specific volume, i.e. the volume of unit mass of matter (so that $1/V$ is the density of the substance).

Let us consider the phase diagram in the V/T plane, and a gas whose specific volume and temperature correspond to some point a in Fig. 82. If the gas is compressed at constant tempera-

ture, then the point representing the state of the gas will move to the left along a straight line parallel to the V axis. At a certain pressure corresponding to the specific volume V_g (the point A), the gas begins to condense into a liquid. As the system is compressed further, the quantity of liquid increases and the quantity of gas decreases; finally, when a certain point B is reached the substance is entirely liquid and its specific volume is V_l .

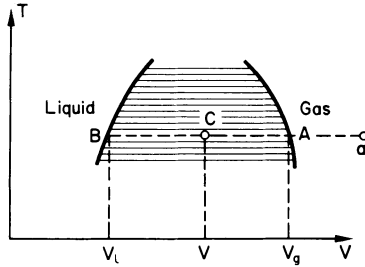


FIG. 82.

The specific volumes of the gas and the liquid which are formed from each other (V_g and V_l) are functions of the temperature at which the transition occurs. When these two functions are represented by appropriate curves, we obtain a phase diagram of the kind shown in Fig. 82. The regions of the diagram to the right and left of the hatched area correspond to the gaseous and liquid phases. The hatched area between the two curves is the region of separation into two phases. The horizontal hatching is significant: the points A and B at which a horizontal line through a point C in this region meets the boundaries of the region give the specific volumes of the liquid and vapour coexisting at that point.

The different points on AB evidently correspond to equilibrium of the same liquid and vapour in different relative amounts. Let the fractions of vapour and liquid at some point C be x and $1-x$. Then the total volume of the system per unit mass is

$$V = xV_g + (1-x)V_l,$$

whence

$$x = \frac{V - V_l}{V_g - V_l}, \quad 1 - x = \frac{V_g - V}{V_g - V_l}.$$

The ratio of these quantities is

$$\frac{x}{1-x} = \frac{V - V_l}{V_g - V} = \frac{BC}{AC}.$$

We see that the quantities of vapour and liquid are inversely proportional to the lengths of AC and BC , i.e. the distances of C from the points A and B which correspond to the pure vapour and pure liquid. This relation is called the *lever rule*.

The phase diagram with pressure instead of temperature as ordinate is exactly similar in appearance. We see that these diagrams do not resemble the diagrams in the pT plane. The region of separation into two phases, which in the pT diagram is only a line, occupies a whole area in the VT and Vp diagrams. This difference arises because phases in equilibrium necessarily have the same temperature and pressure by the general conditions of thermal equilibrium, but their specific volumes are different.

Table 1 shows the melting and boiling points of a number of substances (at atmospheric pressure).

TABLE 1

	Melting point (°C)	Boiling point (°C)
Helium-3	—	-270.0 (3.2°K)
Helium-4	—	-268.9 (4.2°K)
Hydrogen	-259.2 (14°K)	-252.8 (20.4°K)
Oxygen	-219	-183
Ethyl alcohol	-117	78.5
Ethyl ether	-116	34.5
Mercury	- 38.9	356.6
Lead	327	1750
Aluminium	660	2330
Sodium chloride	804	1413
Silver	961	2193
Copper	1083	2582
Iron	1535	2800
Quartz	1728	2230
Platinum	1769	4000
Tungsten	3380	6000

Helium liquefies at a lower temperature than any other substance existing in Nature; the solidification of helium will be discussed in §72. Tungsten has higher melting and boiling points than those of any other chemical element.

§67. The Clausius-Clapeyron equation

The transition of matter from one phase to another always involves the gain or loss of a certain quantity of heat called the *latent heat* or *heat of transition*. When a liquid becomes a gas this is the heat of evaporation; when a solid becomes a liquid, it is the heat of fusion.

Since a phase transition occurs at constant pressure, the heat of transition q_{12} from phase 1 to phase 2 is equal to the difference of the enthalpies W_1 and W_2 of the substance in the two phases (see §56):

$$q_{12} = W_2 - W_1.$$

It is clear that $q_{12} = -q_{21}$, i.e. if heat is absorbed in a given phase transition, the reverse transition is accompanied by the evolution of heat.

In melting and in evaporation, heat is absorbed. These are particular cases of a general rule according to which a phase transition brought about by heating is always accompanied by the absorption of heat. This rule in turn is a consequence of Le Chatelier's principle: heating tends to cause processes to occur which are accompanied by absorption of heat and which therefore as it were act against the external interaction.

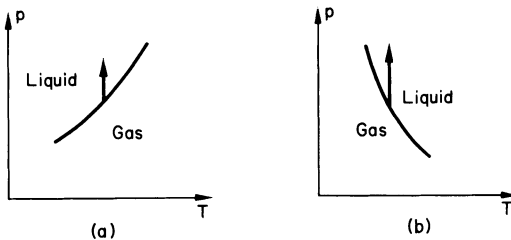


FIG. 83.

The same principle can be used to relate the direction of the phase-equilibrium curve in the pT plane to the change in volume in a phase transition. Let us consider, for example, an equilibrium system consisting of a liquid and a vapour and suppose that it is compressed, so that the pressure in it increases. Then processes must occur in the system which reduce the volume of the substance and thus counteract the effect of the compression. For

this to be so, condensation of the vapour must occur, since the conversion of vapour into liquid is always accompanied by a decrease in volume. This means that, as we move upwards from the equilibrium curve (Fig. 83), we must enter the region of the liquid phase. In this case the liquid is also the “low-temperature” phase, i.e. the phase which exists at lower temperatures. Thus it follows that the equilibrium curve for a liquid and a gas must have the form shown in Fig. 83a, and not that in Fig. 83b; the transition temperature must increase with increasing pressure.

The same relation between transition temperature and pressure must evidently occur whenever the transition to the “high-temperature” phase is accompanied by an increase in volume. For example, since in almost all cases the volume of a substance increases on melting, the melting point usually rises with increasing pressure. In some substances, however, melting is accompanied by a decrease in volume (as for instance in ice, cast iron, and bismuth). For these substances the melting point is lowered by increasing the pressure.

All these qualitative results are expressed quantitatively by a formula which relates the slope of the phase-equilibrium curve, the heat of transition, and the change in volume in the transition.

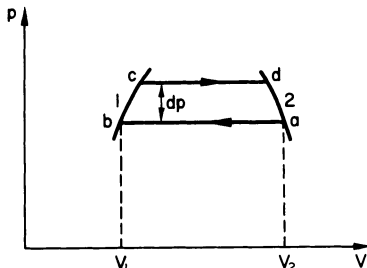


FIG. 84.

To derive this formula, let us imagine a very “narrow” Carnot cycle applied to a certain quantity of substance, the isothermal processes being a transition of the substance from phase 2 to phase 1 at a pressure p and the reverse transition from phase 1 to phase 2 at a pressure $p + dp$. These transitions are represented in the pV phase diagram (Fig. 84) by the lines ab and cd . The sides bc and da should, strictly speaking, be taken as segments of adiabatics, but in the limit of an infinitely narrow cycle the

difference is unimportant and does not affect the area of the cycle, which is the work done in the cyclic process; this work is evidently just $(V_2 - V_1) dp$. The work must also be equal to the product of the quantity of heat q_{12} which is expended (on the isotherm cd) and the efficiency of the Carnot cycle. The quantity q_{12} is simply the heat of transition from phase 1 to phase 2, and the efficiency is dT/T , where dT is the temperature difference between the two isotherms. Thus we have

$$(V_2 - V_1) dp = q_{12} dT/T$$

or

$$\frac{dp}{dT} = \frac{q_{12}}{T(V_2 - V_1)}.$$

This formula, which determines the slope of the phase-equilibrium curve $p = p(T)$, is called the *Clausius-Clapeyron equation*. It may also be written in the form

$$\frac{dT}{dp} = \frac{T(V_2 - V_1)}{q_{12}},$$

where the temperature of the transition is regarded as a function of pressure. In these formulae the volumes V_1 , V_2 of the two phases and the heat q_{12} relate to a given quantity of the substance (e.g. one gram or one gram-molecule).

It should be noted that the derivative dp/dT is inversely proportional to the difference in volume $V_2 - V_1$. Since the change in volume in evaporation is large and that in melting is small, melting curves are much steeper than evaporation curves. For example, to lower the boiling point of water by 1° it is sufficient to reduce the pressure by 27 mm Hg, whereas the same change in the melting point of ice would require the pressure to be increased by 130 atm.

§68. Evaporation

A vapour in equilibrium with its liquid is said to be *saturated*, and its pressure is called the *saturated vapour pressure*. The liquid-vapour equilibrium curve (Fig. 81, §66) may also be

regarded as showing the relation between this pressure and the temperature.

The saturated vapour pressure always increases with increasing temperature. We have seen above that this behaviour is due to the increase in volume of a substance on evaporation. This increase is usually very large. For example, the volume of water vapour at 100°C is 1600 times the volume of water; the boiling of liquid oxygen at -183°C is accompanied by a volume increase by a factor of about 300.

At sufficiently low temperatures the density of the saturated vapour becomes so small that it behaves as an ideal gas. A simple formula can then be derived for the temperature dependence of the vapour pressure. To do so, we use the Clausius-Clapeyron equation,

$$\frac{dp}{dT} = \frac{q}{T(V_g - V_l)},$$

with q the molar heat of evaporation, and V_g and V_l the molar volumes of the vapour and the liquid. Since the volume V_g is very large in comparison with V_l , the latter may be neglected. The volume of one gram-molecule of gas is $V_g = RT/p$. We have

$$\frac{dp}{dT} = \frac{pq}{RT^2},$$

or

$$\frac{1}{p} \frac{dp}{dT} = \frac{d \log_e p}{dT} = \frac{q}{RT^2}.$$

Although the heat of evaporation is itself a function of temperature, it may often be regarded as practically constant over considerable ranges of temperature; for example, the heat of evaporation of water decreases by only 10% between 0 and 100°C. The above formula may then be written as

$$\frac{d \log_e p}{dT} = -\frac{d}{dT} \left(\frac{q}{RT} \right),$$

whence

$$\log_e p = \text{constant} - q/RT$$

and finally

$$p = ce^{-q/RT},$$

where c is a constant coefficient. According to this formula the saturated vapour pressure increases very rapidly (exponentially) with temperature.

The origin of this exponential dependence may be understood as follows. The molecules in a liquid are held together by cohesion forces; to overcome these forces and transfer a given molecule from the liquid to the vapour, work must be done. We may say that the potential energy of a molecule in the liquid is less than its potential energy in the vapour by an amount equal to the heat of evaporation per molecule. If q is the molar heat of evaporation, this difference of potential energies is q/N_0 , where N_0 is Avogadro's number.

We can now use Boltzmann's formula (§54) to show that the increase in the potential energy of a molecule by q/N_0 decreases the gas density by a factor $e^{-q/N_0kT} = e^{-q/RT}$ in comparison with the liquid. The pressure of the vapour is proportional to this expression.

The following are the values of the heats of evaporation and heats of fusion for various substances at atmospheric pressure, in joules per mole:

	q_{ev}	q_{fu}
Helium	80	—
Water	40 500	5 980
Oxygen	6 800	442
Ethyl alcohol	39 000	4 800
Ethyl ether	59 000	7 500
Mercury	28 000	2 350

[It may be noted that from the heat of evaporation (far from the critical point; see §69) we can estimate the magnitude of the van der Waals forces between the molecules. As has been mentioned in §39, it is these forces which bring about the condensation of a substance. Thus, on dividing q_{ev} by Avogadro's number to obtain the heat of evaporation per molecule, we derive a quantity which is a measure of the depth of the minimum on the curve of

van der Waals interaction. The quantity obtained in this way for helium is about one-hundredth of an electron-volt, and for the other liquids in the table it is between one-tenth and a few tenths of an electron-volt.]

In ordinary conditions there is present over the surface of a liquid not only its own vapour but also another gas, namely air. This has little effect on the phase equilibrium; evaporation continues until the partial pressure of the vapour becomes equal to the saturated vapour pressure at the temperature of the liquid.

The presence of the atmosphere considerably affects the evaporation process, however, which has a completely different form according as the saturated vapour pressure at a given temperature is less than or greater than the total pressure on the liquid.

In the former case, the liquid evaporates comparatively slowly from its surface. It is true that the partial pressure of the vapour just above the surface almost immediately becomes equal to the saturated vapour pressure, but this saturated vapour penetrates into the surrounding space only slowly (by diffusion), and further liquid evaporates only as this vapour mixes with the air. The rate of evaporation is, of course, increased by artificially removing the vapour from the surface of the liquid.

The process is different when the saturated vapour pressure becomes equal to or slightly greater than the ambient pressure and the liquid *boils* violently. This is shown by the intensive formation, on the surface of the vessel, of gas bubbles which grow by the evaporation of liquid into the bubbles and then become detached and rise through the liquid, causing mixing of it; a stream of vapour passes from the free surface of the liquid into the surrounding medium.

For reasons which will be discussed later (see §99), the conversion of a liquid into a vapour cannot in general occur by spontaneous generation of vapour bubbles within a pure liquid. The centres of formation of the gas phase are tiny bubbles of other gases which already exist on the vessel walls or are formed thereon (or on particles suspended in the liquid) from gases dissolved in the liquid which are expelled on heating. Until the boiling point is reached (at which the saturated vapour pressure becomes equal to the external pressure), the pressure of the surrounding liquid prevents the growth of these bubbles.

By careful previous purification and degassing of the liquid and the vessel walls it is possible to eliminate practically all *vaporisation centres* in it (as may also happen during the boiling process itself). This leads to *superheating*, the liquid remaining a liquid at temperatures above the boiling point. On the other hand, in order to avoid superheating and to ensure that boiling occurs, various artificial sources of vaporisation centres are placed in the vessel of liquid (porous objects, pieces of glass capillary, and so on).

A superheated liquid (i.e. a liquid at a temperature at which it would be expected to have become a gas, at the pressure concerned) is an example of what are called *metastable* states. These are states of limited stability. Although they can exist (when suitable precautions are taken) for a longer or shorter time, the equilibrium is relatively easily destroyed and the substance enters a different state, which is stable. For example, a superheated liquid boils instantaneously when vaporisation centres are created in it.

Similar phenomena occur in the reverse process of condensation of a vapour. Here again the occurrence of the phase transition, in the absence of liquid in contact with the vapour, requires the existence of *condensation centres* in the vapour, usually in the form of small impurities, as will be further discussed in §99. For this reason *supercooling* or *supersaturation* of a vapour is possible, in which it is brought into a state where the pressure exceeds the saturated vapour pressure at the temperature concerned. Such states can be reached, for example, by cooling a carefully purified saturated vapour by adiabatic expansion.

§69. The critical point

As the temperature rises, the saturated vapour pressure increases rapidly, and so does the density of the vapour, approaching that of the liquid. At a certain temperature the density of the vapour becomes equal to that of the liquid, and the vapour and liquid become indistinguishable. In other words, the equilibrium curve of the liquid and gas in the pT phase diagram terminates at some point (K in Fig. 85). This is called the *critical point*, and its coordinates are called the *critical temperature* T_c and *critical pressure* p_c of the substance.

In the VT diagram (and similarly in the Vp diagram) the approach to the critical point is shown by the approach to equality

of the specific volumes of the liquid and vapour as the temperature increases, i.e. by the approach of the two curves which form the boundaries of the hatched region in Fig. 82 (§66). For $T = T_c$ the two curves join, and we thus have essentially a single smooth curve with a maximum at K (Fig. 86). This is the critical point, its coordinates being the critical temperature T_c and the critical specific volume V_c .

As the properties of the liquid approach those of the gas, the heat of transition q between them decreases, and becomes zero at the critical point.

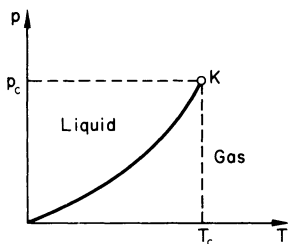


FIG. 85.

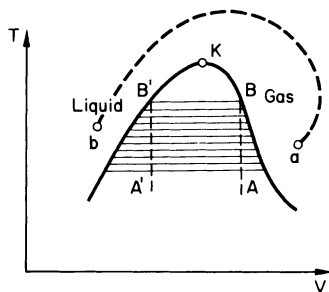


FIG. 86.

The existence of the critical point very clearly demonstrates that there is no fundamental difference between the liquid and gaseous states of matter. For, when considering any two states a and b (Fig. 86) of very different density, we call the denser state b the liquid state, and the less dense a the gaseous state. By compressing the gas a at constant temperature we can convert it to the liquid b , passing through a stage of separation into two phases. But the passage between the same states a and b can also be carried out by first raising and then lowering the temperature while the volume is decreased, in such a way as to move along a path in the VT plane which passes above the critical point, as shown by the broken line in Fig. 86. In this case there is no discontinuous change of state anywhere, the substance remains homogeneous, and we can not say that the substance ceases at some point to be a gas and becomes a liquid.

From the diagram in the VT plane we can easily find what happens on heating a closed vessel (for example, a sealed tube)

containing a certain quantity of liquid and the vapour above it. Since the total volume of the substance is constant, this will correspond to movement upward along a vertical line in the VT plane. If the volume of the tube exceeds the critical volume corresponding to the given quantity of substance, this line will lie to the right of the critical point (AB in Fig. 86) and as heating proceeds the quantity of liquid will decrease until the whole of the substance is converted into vapour (at B); the boundary (meniscus) between the liquid and the vapour will disappear at the lower end of the tube. If the volume of the tube is less than the critical volume (point A'), vapour will condense on heating until the whole of the substance becomes liquid (at B'); the meniscus will disappear at the top of the tube. Finally, if the volume of the tube is equal to the critical volume, the meniscus will disappear somewhere within the tube at the critical temperature T_c .

The values of the absolute critical temperatures T_c , pressures p_c and densities ρ_c for a number of substances are shown in Table 2.

TABLE 2

	T_c (°K)	p_c (atm)	ρ_c (g/cm ³)
Water	647.2	218.5	0.324
Alcohol	516.6	63.1	0.28
Ether	467.0	35.5	0.26
Carbon dioxide	304.2	73.0	0.46
Oxygen	154.4	49.7	0.43
Hydrogen	33.2	12.8	0.031
Helium-4	5.25	2.26	0.069
Helium-3	3.33	1.15	0.041

It has already been mentioned in §52 that solid (crystalline) substances differ fundamentally from liquids and gases in being anisotropic. The transition between a liquid and a crystal therefore cannot be made in a continuous manner as can that between a liquid and a gas. We can always say to which of the two phases (crystal or liquid) a body belongs, according to whether it does or does not have the qualitative property of anisotropy. Thus there cannot exist a critical point for melting.

§70. Van der Waals' equation

As the density of a gas increases, its properties deviate more and more from those of an ideal gas, and it finally condenses to a liquid. These phenomena depend on complex molecular interactions, and there is no way of giving a quantitative description of these in order to derive theoretically an exact equation of state for the substance. We can, however, construct an equation of state which takes account of the main qualitative properties of molecular interaction.

The nature of the interaction between molecules has already been described in §39. The repulsive forces which rapidly increase at short distances signify, roughly speaking, that the molecules occupy a certain definite volume, and the gas cannot be compressed beyond this. Another fundamental property of the interaction is that there is attraction at large distances; this attraction is very important, since it is responsible for the condensation of a gas into a liquid.

First of all, let us take into account, in the equation of state (which will be written for one mole of substance), the limited compressibility of the gas. To do so, we must replace the volume V in the ideal-gas equation $p = RT/V$ by $V - b$, where b is some positive constant which takes into account the size of the molecules. The equation

$$p = RT/(V - b)$$

shows that the volume cannot be made less than b , since for this value of V the pressure becomes infinite.

Let us now take into account the attraction between molecules. This attraction must cause a decrease in the gas pressure, since each molecule near the wall of the vessel is subject to a force towards the interior of the vessel exerted by the other molecules. As a rough approximation, this force on each molecule may be taken as proportional to the number of molecules per unit volume, i.e. to the density of the gas. The pressure itself is also proportional to this number. Thus the total decrease in pressure due to the attraction between molecules is proportional to the square of the gas density, i.e. inversely proportional to the square of its volume. Accordingly we subtract from the above expression for the pressure a term a/V^2 , where a is another constant rep-

resenting the forces of molecular attraction. Thus we have the equation

$$p = \frac{RT}{V-b} - \frac{a}{V^2},$$

or

$$\left(p + \frac{a}{V^2}\right)(V-b) = RT.$$

This is *van der Waals' equation*. When the gas density is low, i.e. the volume V is large, a and b may be neglected, and we return to the equation of state of an ideal gas. We shall see below that the same equation correctly describes the phenomena which occur in the opposite limiting case of high compression.

To examine the behaviour of a gas described by van der Waals' equation, let us consider the isotherms defined by this equation, i.e. the curves of p as a function of V for given values of T . For this purpose we write the equation in the form

$$V^3 - \left(b + \frac{RT}{p}\right)V^2 + \frac{a}{p}V - \frac{ab}{p} = 0.$$

For given values of p and T this is a cubic equation in V .

A cubic equation has three roots, of which either all three or one may be real; in the latter case the equation also has two complex conjugate roots. The volume can, of course, be represented as a physical quantity only by real (and positive) roots. In the present case the equation cannot have negative roots (if p is positive), since if V is negative every term in the equation is negative and their sum cannot be zero. Thus we see that according to van der Waals' equation there are either three different values or one value of the volume corresponding to given values of the temperature and pressure.

The second case always occurs at sufficiently high temperatures. The corresponding isotherms differ from those of an ideal gas only by some change in shape, but remain monotonically decreasing (curves 1 and 2 in Fig. 87; increasing numbers on the curves correspond to decreasing temperatures). At lower temperatures the isotherms have a maximum and a minimum

(curves 4, 5, 6), and so for each of them there is a range of pressures in which the curve gives three values of V (three points of intersection of the isotherm with a horizontal line).

Figure 88 shows one such isotherm. Let us see what is the significance of its various parts. On the sections ge and ca the dependence of pressure on volume is of the normal type: the pressure increases as the volume decreases. The section ec would correspond to the unnatural situation where compression of the substance would decrease the pressure. It is

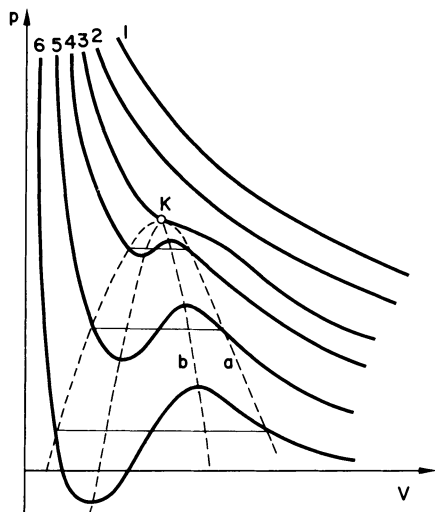


FIG. 87.

easily seen that such states cannot exist in Nature. For let us imagine a substance with these properties, and suppose that a small region of it happens to contract, e.g. owing to the fluctuations described in §64. Then its pressure will decrease also, i.e. become less than the pressure of the surrounding medium, which in turn causes a further contraction, and so on, i.e. this small region will contract at an increasing rate. Thus these states of matter would be completely unstable and therefore could not occur in reality.

The existence of the unrealisable section ec of the isotherm signifies that, as the volume gradually varies, the substance cannot remain homogeneous at all times: at some point there

must be a discontinuous change of state and the substance must separate into two phases. In other words, the true isotherm is given by the curve *abfg*. The part *ab* corresponds to the gaseous state of the substance, and *fg* to the liquid state. The straight horizontal section *bf* corresponds to two-phase states where the gas becomes a liquid; this occurs at a certain constant pressure (for a given temperature). [It can be shown that the section *bcd* must be situated so that the areas *bcd* and *def* are equal.]

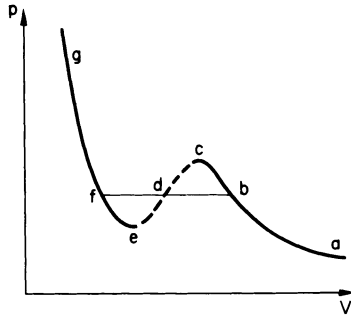


FIG. 88.

The sections *bc* and *ef* of the isotherm correspond to metastable states of supercooled vapour and superheated liquid (§68). We now see that there are limits (represented by the points *c* and *e*) beyond which the vapour cannot be supercooled or the liquid superheated.

As the temperature rises, the straight section of the isotherm becomes shorter, and at the critical temperature it contracts to a point (*K* in Fig. 87). The isotherm 3 which passes through this point separates isotherms of two types: the monotonic isotherms (1,2) and the isotherms with minima and maxima (4,5,6), on which the substance must necessarily separate into two phases.

If the beginning and end of the straight section of each isotherm are joined by a curve (*a* in Fig. 87), this gives the curve of phase equilibrium of liquid and vapour in the *pV* diagram. The maximum *K* on this curve is the critical point. On joining the points which correspond to *c* and *e* in Fig. 88 we obtain a curve (*b* in Fig. 87) which is the boundary of the region within which the substance cannot exist without separation into two phases, even in a metastable state.

At the critical point the three points at which the straight section intersects the van der Waals isotherm merge into one. Hence it follows that the tangent to the isotherm at the critical point is horizontal, i.e. the derivative of the pressure with respect to the volume (at constant temperature) is zero: $(dp/dV)_T = 0$. The reciprocal of this quantity is the compressibility of the substance, which is therefore infinite at the critical point.

The section of the isotherm which corresponds to a superheated liquid may lie partly below the axis of abscissae (as on the isotherm 6 in Fig. 87). This section corresponds to metastable states of an "expanded" liquid, as discussed at the end of §51.

§71. The law of corresponding states

The critical values of the volume, temperature and pressure can be related to the parameters a and b in van der Waals' equation. To do so, we note that for $T = T_c$ and $p = p_c$ all three roots of van der Waals' equation

$$V^3 - \left(b + \frac{RT_c}{p_c}\right)V^2 + \frac{a}{p_c}V - \frac{ab}{p_c} = 0$$

are the same and equal to the critical volume V_c . This equation must therefore be identical with

$$(V - V_c)^3 = V^3 - 3V^2V_c + 3VV_c^2 - V_c^3 = 0.$$

A comparison of coefficients of powers of V in the two equations gives the three relations

$$b + \frac{RT_c}{p_c} = 3V_c, \quad \frac{a}{p_c} = 3V_c^2, \quad \frac{ab}{p_c} = V_c^3.$$

These relations, regarded as equations for the unknowns V_c , p_c and T_c , are easily solved to give

$$V_c = 3b, \quad p_c = a/27b^2, \quad T_c = 8a/27bR.$$

By means of these relations we can carry out the following interesting transformation of van der Waals' equation. In this equation we write, instead of the variables p , T , V , their ratios to the critical values:

$$p^* = p/p_c, \quad T^* = T/T_c, \quad V^* = V/V_c;$$

these ratios are called the *reduced* pressure, temperature and volume. By means of simple transformations we can easily see that van der Waals' equation then becomes

$$\left(p^* + \frac{3}{V^{*2}}\right)(3V^* - 1) = 8T^*.$$

The precise form of this equation is not of particular interest; the remarkable thing is that it does not involve the constants a and b which depend on the nature of the gas. In other words, if the critical values are used as the units of measurement of the volume, pressure and temperature, the equation of state becomes the same for all substances. This is called the *law of corresponding states*.

If this law applies to the equation of state, it will apply also to all phenomena which are in any way related to the equation of state, including the gas-liquid phase transition. For example, the temperature dependence of the saturated vapour pressure, if written as a relation between reduced quantities $p/p_c = f(T/T_c)$, must be a universal relation.

A similar conclusion may be drawn concerning the heat of evaporation q . Here we must consider the dimensionless ratio of the heat of evaporation to some other quantity of the same dimensions (energy/mole); this may be taken as RT_c . According to the law of corresponding states, the ratio q/RT_c must be the same for all substances as a function of the reduced temperature: $q/RT_c = F(T/T_c)$. For temperatures much below the critical temperature this function tends to a constant limit, whose experimental value is about 10.

It should be emphasised that the law of corresponding states is only approximate, but it can be used to derive results which are entirely suitable for rough estimations.

Although the law of corresponding states has been derived from van der Waals' equation, it is in fact somewhat more accurate than the latter, since it does not depend on the specific form of the equation of state, but follows simply from the fact that this equation involves only two constants a and b . A different equation of state with two parameters would likewise lead to the law of corresponding states.

§72. The triple point

As we know, equilibrium between two phases is possible only when a certain relation holds between the temperature and the pressure, represented by a certain curve in the pT plane. It is evident that three phases of the same substance cannot be simultaneously in equilibrium with one another along a line; such an equilibrium is possible only at a particular point in the pT diagram, i.e. at a particular pressure and a particular temperature. This is the point at which the equilibrium curves of each pair of the three phases intersect. Points of equilibrium of three phases are called *triple points*. For example, for water the simultaneous existence of ice, steam and liquid water is possible only at 4.62 mm Hg pressure and $+0.01^{\circ}\text{C}$ temperature.

Since even three phases are in equilibrium only at one point, four or more phases cannot exist simultaneously in equilibrium with one another.

The fact that triple points correspond to definite values of the temperature makes them especially suitable as fixed points of the temperature scale. Their reproduction is free from the difficulties associated with the need to maintain a given pressure, as is required, for example, when the melting point of ice at atmospheric pressure (or any point of equilibrium of two phases) is taken as a fixed point. The precise definition now used for the absolute degree is based on such a choice: the temperature of the triple point of water is taken to be exactly 273.16°K . It should be mentioned, however, that with the present accuracy in the measurement of temperature and pressure this definition is indistinguishable from that in which the melting point of ice is taken as 273.15°K .

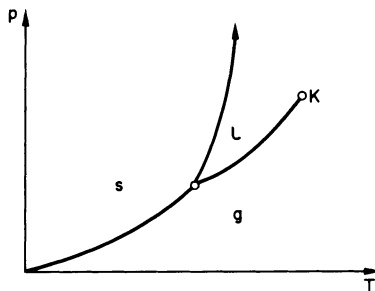


FIG. 89.

Figure 89 shows the form of the phase diagram for a substance having only three phases: solid, liquid and gaseous. In the diagram, the regions marked *s*, *l* and *g* correspond to these phases, and the lines separating the regions are the curves of equilibrium of the corresponding pairs of phases. The direction of the melting curve is chosen so as to correspond to the usual case where a body expands on melting (see §67). For the few instances where melting is accompanied by contraction of the substance, the curve slopes in the opposite direction.

It is seen from the phase diagram that a substance does not necessarily pass through a liquid state in the course of becoming a gas. At pressures below the triple point, heating the solid converts it directly into vapour; this phase transition is called *sublimation*. For example, solid carbon dioxide sublimates at atmospheric pressure, since its triple point corresponds to a pressure of 5.1 atm (and a temperature of -56.6°C).

The curve of equilibrium of a liquid and a gas terminates at the critical point (*K* in Fig. 89). For transitions between liquid and solid phases there can be no critical point (as already mentioned in §69). The melting curve therefore cannot simply terminate, and must continue indefinitely.

The curve of equilibrium of a solid and a gas passes through the origin, i.e. at absolute zero temperature a substance is in the solid state at any pressure. This is a necessary consequence of the ordinary concept of temperature based on classical mechanics. According to this concept, the kinetic energy of the atoms is zero at absolute zero temperature, i.e. all the atoms are at rest. The equilibrium state of a body is then one in which the configuration of the atoms corresponds to the minimum energy of interaction between the atoms. This configuration, whose properties differ from all others, must have some degree of order-

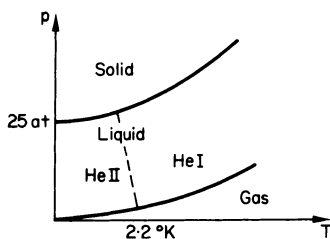


FIG. 90.

ing, i.e., must represent a spatial lattice. This means that the substance must be crystalline at absolute zero.

There exists in Nature, however, one exception to this rule: helium, after becoming liquid, remains liquid at all temperatures down to absolute zero. The phase diagram of the isotope helium-4 is shown in Fig. 90; the broken line in this diagram will be explained in §74. We see that the evaporation and melting curves nowhere intersect, i.e. there is no triple point. The melting curve meets the ordinate axis at $p = 25$ atm; this means that, in order to solidify helium, it must be not only cooled but at the same time subjected to a pressure of at least 25 atm.

It is clear from the above that this behaviour of helium is inexplicable on the basis of classical ideas; it is in fact due to quantum effects. As already mentioned in §50, according to quantum mechanics the motion of the atoms does not cease entirely even at absolute zero. For this reason the above conclusion that a substance must solidify at this temperature is also incorrect. The quantum properties of a substance appear more markedly at low temperatures, where they are not masked by the thermal motion of the atoms. All substances except helium solidify before their quantum properties become sufficiently important; only helium becomes a "quantum liquid" which need not solidify. Other remarkable properties of this liquid will be discussed in §124.

§73. Crystal modifications

The region of the solid state is not usually occupied by a single phase. At different pressures and temperatures a substance may be in different crystal states, each with a definite structure. These different states are also different phases of the substance, and are called *crystal modifications*; the property of having more than one such modification is very common, and is called *polymorphism* of the substance (or, for the elements, *allotropy*). Well-known examples are the modifications of carbon (graphite and diamond), sulphur (which forms orthorhombic and monoclinic crystals), and silica (the various minerals quartz, tridymite, cristobalite).

Like any other phases, different modifications can be in equilibrium with one another only along certain lines in the pT diagram; the transition from one modification to another (called a poly-

morphic transformation) is accompanied by absorption or evolution of heat. For example, the transformation of what is called α iron (with a body-centred cubic lattice) into γ iron (with a face-centred cubic lattice) occurs at 910°C at atmospheric pressure, and is accompanied by the absorption of about 1600 J/mole of heat.

Figure 91 shows, as an example, the general form of the phase diagram for sulphur. The letters *R* and *M* denote the regions where the two solid phases are stable: rhombic (ordinary yellow sulphur) and monoclinic. We see that here there are three triple points.

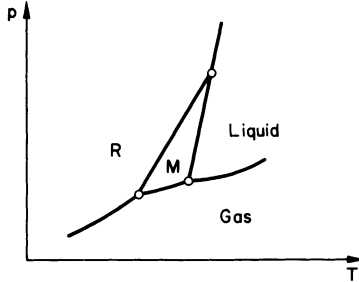


FIG. 91.

Figure 92 shows the phase diagram of water. The five crystal modifications of ice are shown in the diagram by the numbers I, II, III, V, VI. Ordinary ice corresponds to region I; the other modifications are formed only under pressures of thousands of atmospheres. The vapour region corresponds to such low pressures that it is almost impossible to show it in the same diagram.

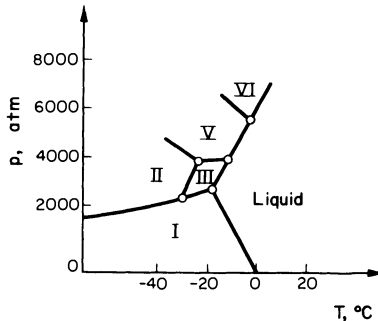


FIG. 92.

A typical feature of phase transitions between different crystal modifications is the ease with which metastable states can occur. Supercooling of a vapour or superheating of a liquid is possible only when the necessary precautions are taken, but the delay of phase transitions in the solid state and existence of crystal modifications in conditions where they are “not permitted” are almost the rule. This is quite understandable, since the closeness of atoms in a crystal and the restriction of their thermal motion to small oscillations greatly hamper the rearrangement of the lattice into a different modification. An increase in temperature makes the thermal motion more violent and thus accelerates this rearrangement.

Here it should be remembered that the polycrystalline structure of a solid is itself in a sense metastable (in comparison with the single-crystal state). Thus, when a body composed of small crystals is heated, its component crystals become larger, some crystals growing at the expense of others; this is called *recrystallisation*. The amorphous state of a body may also be metastable; for example, spontaneous crystallisation is the reason for the cloudiness of very old glass.

A polymorphic transformation is facilitated by the presence in the former phase of inclusions of the new phase, which act as “nuclei”. A well-known example of this is the transformation of ordinary white tin (which has a tetragonal structure) into a powder of grey tin (a modification which has a cubic lattice). At atmospheric pressure these modifications are in equilibrium at 18°C, white tin being stable above this temperature and grey tin below. In practice, however, white tin can exist even below freezing point, but when a few grains of the grey modification are added it crumbles to a grey powder.

The difficulty of lattice rearrangement at low temperatures may bring about the existence of modifications which are not stable phases under any conditions; such modifications do not appear at all in the phase diagram, which represents stable states of a substance. This is observed, for example, in the hardening of steel. The solid solution of carbon in γ iron, called austenite, is stable only at temperatures 700–900°C (depending on the carbon content), and must decompose at lower temperatures. When austenite is very rapidly cooled or *quenched*, however, there occurs instead a formation in the metal of needle-shaped

crystals of a new phase, a solid solution with a tetragonal lattice called martensite, which is extremely hard. This “intermediate” phase is always metastable, and decomposes when steel is slowly heated or *tempered* at 250–300°.

Figure 93 shows the phase diagram of carbon; the region of the gaseous state is at low pressures and is not visible on the scale used in this diagram. The phase diagram shows that at ordinary pressures and temperatures the stable modification is graphite. But under ordinary conditions graphite and diamond both exist as almost completely stable crystals. This is due to

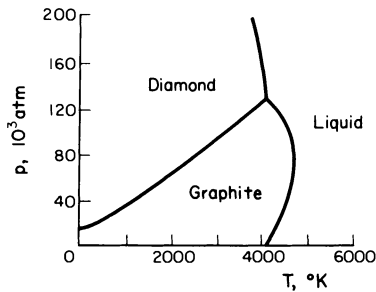


FIG. 93.

the large difference in the structures of the two crystals, which requires a very extensive rearrangement to convert one into the other (as is indicated by the fact that the density of diamond is 1.5 times that of graphite). On heating to high temperatures, however, diamond is transformed into graphite: above 1700°K it rapidly crumbles into graphite powder (if heated in a vacuum to prevent combustion). The diagram shows that the reverse process of conversion of graphite into diamond can occur only at very high pressures. The region in which diamond is stable lies above about 10 000 atm. A high temperature is also necessary if the process is to occur at a reasonable rate. The process is carried out in practice at pressures of 50 000 to 100 000 atm and temperatures of 1500–3000°K, and a metal catalyst must also be present. The spontaneous transformation of graphite into diamond has been observed at about 130 000 atm at temperatures above 3300°K; this is apparently in the region where graphite is neither stable nor metastable, but completely unstable.

§74. Phase transitions of the second kind

It has already been mentioned that the transition between phases of different symmetry cannot occur in crystals in a continuous manner as it can in a liquid and a gas. In every state the body has one symmetry or the other, and therefore we can always assign it to one of the two phases.

The transition between different crystal modifications is usually effected by means of a phase transition in which there is a sudden rearrangement of the crystal lattice and the state of the body changes discontinuously. As well as such discontinuous transitions, however, another type of transition involving a change of symmetry is also possible.

Actual examples of such transitions are somewhat complicated as regards the details of the crystal structure of the bodies. To illustrate the nature of these transitions, we shall therefore consider an imaginary example. Let us suppose that a body crystallises at low temperatures in the tetragonal system, i.e. has a lattice consisting of cells which are rectangular parallelepipeds with square bases and a height c greater than the side a of the base. Let the difference between a and c be small, i.e. the crystal be tetragonal but almost cubic; and let us assume that during thermal expansion the a edges increase in length more rapidly than the height c . Then, as the temperature increases, the lengths of the sides of the unit parallelepiped will become more nearly equal and at a certain temperature they will be the same; on further heating, all three sides will increase in length at the same rate, remaining equal. It is clear that, as soon as a becomes equal to c , the symmetry of the lattice suddenly changes from tetragonal to cubic, and we have essentially a different modification of the substance.

This example is typical in that there is no discontinuous change in state of the body. The configuration of the atoms in the crystal changes continuously. However, an arbitrarily small displacement of the atoms from their symmetrical position in the lattice of the cubic modification (when the temperature falls again) is sufficient to cause a sudden change in the symmetry of the lattice. So long as all three sides of the cell are equal the lattice is cubic, but the appearance of even an infinitesimal difference between the lengths a and c makes the lattice tetragonal.

A transition between crystal modifications which occurs in

this way is called a *phase transition of the second kind*, in contrast to ordinary phase transitions, which in this case are said to be of the first kind. †

Thus a phase transition of the second kind is continuous in the sense that the state of the body changes continuously. It should be emphasised, however, that the symmetry at the transition point does, of course, change discontinuously, so that we can always assign the body to one of the two phases. But, whereas at a phase transition point of the first kind bodies in two different states are in equilibrium, at a transition point of the second kind the states of the two phases are the same.

The absence of a discontinuity of state in a phase transition of the second kind means that there is no discontinuity in quantities which describe the thermal state of a body: volume, internal energy, enthalpy, etc. Hence, in particular, such a transition is not accompanied by evolution or absorption of heat.

Nevertheless, at a transition point there is a discontinuous change in the dependence of these quantities on temperature. For instance, in the example considered it is evident that the thermal expansion of the crystal will occur differently according as there is only a change in the volume of the lattice (when the crystal has cubic symmetry) or the heating also brings about a change in cell shape as a result of unequal changes in the height and base edge of the cells, as when there is tetragonal symmetry. It is also evident that different quantities of heat will be necessary for the same temperature increase.

This means that at a transition point of the second kind there is a discontinuity in the temperature derivatives of the thermal properties of the body, i.e. in the coefficient of thermal expansion $(dV/dT)_p$, the specific heat $C_p = (dW/dT)_p$, etc.

The presence of these discontinuities is the main characteristic of transitions of the second kind which appears in thermal measurements. Figure 94 shows the typical manner of variation of specific heat with temperature near a transition point of this

† The example described above is not entirely imaginary. A change of this type occurs in the lattice of barium titanate (BaTiO_3). At room temperature this lattice is tetragonal, with values of a and c which differ by 1%. When the temperature increases, the length a increases and c decreases. At 120°C the substance changes to the cubic modification, but in this actual case the values of a and c do in fact have a slight discontinuity at the transition point, so that the transition is of the first kind.

kind: a gradual increase is interrupted by a sudden drop, after which the specific heat again begins to rise.

In a transition of the second kind, the pressure derivatives of thermal quantities are also discontinuous. For example, the compressibility $(dV/dp)_T$ has a discontinuity.

Let us return to the imaginary example used above. The following property of the change in symmetry in the transition may be noted: the lattice of the cubic modification has all the symmetry of the elements of the tetragonal modification and some other elements as well. In this sense we can say that the transition takes place between two phases of which one has higher symmetry than the other. This is in fact a general property and applies to all phase transitions of the second kind.

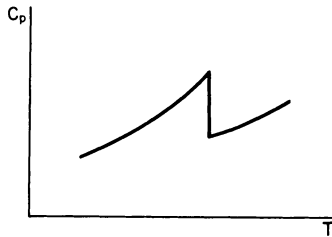


FIG. 94.

This is a limitation (in reality not the only one) on the possible existence of a phase transition of the second kind. For example, no such transition can occur between crystals of the cubic and hexagonal systems: neither of these symmetries can be said to be higher than the other, since the former contains axes of order 4 which do not occur in the latter, but it does not, on the other hand, contain the axis of order 6.

It can also be shown that a transition of the second kind cannot occur between a crystal and a liquid.

The direction in which the specific heat changes discontinuously at a transition of the second kind is related to the way in which the symmetry changes: the specific heat is smaller in the phase of higher symmetry. In most cases the high-temperature phase has the higher symmetry, and the discontinuity of the specific heat is then as shown in Fig. 94. This sequence of the phases with respect to temperature is not necessary, however. For example, Rochelle salt ($\text{NaK}(\text{C}_4\text{H}_4\text{O}_6) \cdot 4\text{H}_2\text{O}$) has two transition

points of the second kind (at -18 and -23°C), between which its crystals belong to the monoclinic system; at temperatures outside this range the salt forms orthorhombic crystals. It is clear that the passage through the upper point in the direction of increasing temperature is accompanied by an increase of symmetry, but the passage through the lower point involves a corresponding decrease of symmetry.

It has already been mentioned that ordinary phase transitions frequently exhibit phenomena of superheating or supercooling, in which one phase continues to exist (as a metastable phase) under conditions where the other phase is stable. The nature of these phenomena depends on the necessity for "centres" on which the new phase can grow. In transitions of the second kind, such phenomena are obviously impossible, since one phase changes into the other instantaneously and continuously. This is very clearly seen in the example considered above, where the transition amounted essentially to a change in the configuration of the atoms in thermal expansion.

Phase transitions of the second kind are not always transitions between different crystal modifications, but they always bring about some new qualitative property of a body, with a continuous change of state. This may be a new symmetry property (related to the magnetic properties of the substance), or it may be the occurrence of what is called superconductivity, the disappearance of electrical resistance.

Finally, there is a very unusual phase transition of the second kind in liquid helium at about 2.2°K . In this transition the liquid remains a liquid but acquires fundamentally new properties (see §124). The broken line in the helium phase diagram (Fig. 90, §72) divides the regions of existence of the two phases, which are known as helium I and helium II.

§75. Ordering of crystals

All the crystal structures discussed in §47 have the property that the atoms of each kind are situated at entirely definite positions and, conversely, at each lattice point there must be an atom of a particular kind. The number of atoms of each kind is equal to the number of places for them in the lattice.

There are also structures which do not have this property, however; for example, that of sodium nitrate (NaNO_3). We shall

not describe this in detail, but simply mention that in this crystal the NO_3 groups form layers in which the nitrogen atoms are at the vertices of equilateral triangles and the oxygen atoms surround the nitrogen atoms in position *a* or *b* (Fig. 95). The possibility of these two orientations of the NO_3 groups implies that the number of positions which can be occupied by oxygen atoms is twice the number of these atoms.

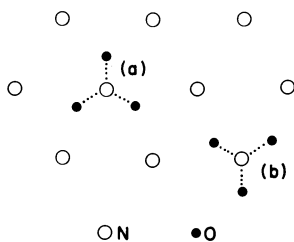


FIG. 95.

At sufficiently low temperatures, the oxygen atoms take up quite definite positions; in practice, what happens is that in each layer all the NO_3 groups have the same orientation, and layers with orientation *a* alternate with those having orientation *b*. Such a crystal is said to be *completely ordered*.

When the temperature is raised, however, the ordered arrangement of the atoms is disturbed: as well as NO_3 groups occupying the usual (their "own") position, there appear groups with the "other" orientation.

As the degree of ordering decreases, i.e. as the fraction of "incorrectly" oriented NO_3 groups increases, a point is finally reached, at a temperature of 275°C , where the "own" and "other" orientations are entirely mixed: each NO_3 group has an equal probability of occupying either position. The crystal is then said to be *disordered*. All the NO_3 layers become crystallographically equivalent, i.e. there is a change, namely an increase, in the symmetry of the crystal.

Phenomena of crystal ordering occur very widely in alloys. For example, crystals of brass (the alloy CuZn) at low temperatures have a cubic lattice with the copper atoms at the vertices and the zinc atoms at the centres of the cubic cells (Fig. 96a). This structure corresponds to a completely ordered crystal.

The copper and zinc atoms may change places, however; in this sense we may say that in this crystal also the number of places available to atoms of each kind is twice the number of those atoms. As the temperature increases, the number of “incorrectly” placed atoms becomes larger, and complete disorder exists at 450°C : at each lattice point a copper atom or a zinc atom can occur with equal probability, so that all the lattice points become equivalent (Fig. 96b). At this stage the symmetry of the crystal obviously changes: its Bravais lattice becomes body-centred cubic instead of simple cubic.

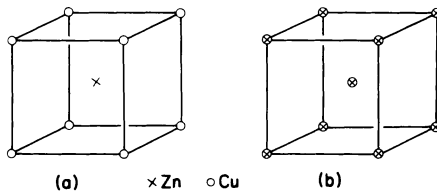


FIG. 96.

In both the examples described above, the transition to the disordered state occurs by a phase transition of the second kind. The degree of ordering decreases continuously and becomes zero at a certain temperature, which is the transition point.

This type of transition to the disordered state is not a general rule, however; the change can also occur by an ordinary discontinuous phase transition. In such cases the ordered configuration of the atoms in the crystal is destroyed at first only to a comparatively small extent as the temperature increases, and at a certain temperature the crystal suddenly enters the disordered state, in which the atoms are completely intermingled. Such a transition occurs, for example, at 390°C in the alloy Cu_3Au . In its disordered phase the copper and gold atoms are randomly located at all the points of a face-centred cubic lattice; in the ordered crystal the gold atoms occupy positions at the vertices of the cubic cells, and the copper atoms at the centres of the faces.

§76. Liquid crystals

In addition to the anisotropic crystalline and isotropic liquid states, a substance may also exist in a peculiar state called a

liquid crystal. In its mechanical properties, a substance in this state resembles an ordinary liquid, being fluid; liquid crystals include substances both of high mobility (low viscosity) and of low mobility (high viscosity). These liquids nevertheless differ from ordinary liquids in being anisotropic; this is most noticeable in their optical properties.

The liquid-crystal state is observed in many complex organic substances having large molecules, usually of elongated form. It is not uncommon; about one complex organic substance out of every two hundred forms liquid crystals.

The physical nature of the liquid-crystal state appears to be as follows. In an ordinary liquid the relative position and orientation of the molecules are completely random; in other words, the molecules of a liquid in their thermal motion undergo both random translational movements and random rotations. In a liquid crystal, however, although the molecules are randomly situated in space, their mutual orientation is not random. In other words, only the translational thermal motion of the molecules is random, and not their rotation. The simplest example of such a structure can be imagined as a liquid consisting of rod-shaped molecules, which can move in any manner relative to one another provided that they remain parallel. Since there is no obstacle to the translational motion of the molecules, the substance is fluid, i.e. behaves as a liquid, but the ordered arrangement of the molecules has the result that the substance is anisotropic. For instance, it is clear that the properties of the substance in the direction parallel to the rod-shaped molecules will be entirely different from its properties in other directions.

A substance in the liquid-crystal state is not usually a "single crystal", but forms a "polycrystalline" mass consisting of a large number of droplet-like liquid crystallites oriented variously with respect to one another. For this reason, a substance which is a liquid crystal usually has the appearance of a turbid liquid: this occurs because of the random scattering of light at the boundaries between droplets. By means of a strong electric or magnetic field it is possible in some cases to give all the droplets the same orientation, and an almost clear liquid "single crystal" is obtained.

If a liquid crystal is placed in a liquid with which it does not mix, the individual liquid-crystal drops take a form which is

sometimes spherical, sometimes ellipsoidal, and in a few cases even that of strange polyhedra with much rounded edges and corners.

Substances which exist in the liquid-crystal state also possess ordinary solid-crystal and isotropic liquid phases. The sequence of formation of these phases is as follows. At low temperatures the substance is a solid crystal, at higher temperatures it enters the liquid-crystal state, and at still higher temperatures it becomes an ordinary liquid. Many substances form not only one but two or more different liquid-crystal modifications. Like all phase transitions, the transformations of liquid-crystal phases into one another or into other phases occur at precisely defined temperatures and are accompanied by the evolution or absorption of heat.

CHAPTER X
SOLUTIONS

§77. **Solubility**

Solutions are mixtures of two or more substances in which the substances are mixed on the molecular scale. The relative amounts of the various substances in the mixture may vary over a more or less wide range. If one substance is present in greater quantity than the others, it is called the *solvent*, and the other substances are called *solutes*.

The composition of a solution is described by its *concentration*, which gives the relation between the quantities of the substances in the mixture – the components of the mixture, as they are called. The concentration can be defined in various ways. Physically, the most informative is the molar concentration, i.e. the ratio of the numbers of molecules (or, what is the same thing, the ratio of the quantities expressed in moles). Alternatively, we may use concentrations by weight, by volume (the volume of substance dissolved in a given volume of solvent), and so on.

The process of dissolution is accompanied by the evolution or absorption of heat. The quantity of heat depends not only on the quantity of solute but also on the quantity of solvent.

The *heat of solution* is usually defined as the quantity of heat evolved or absorbed in the dissolution of one gram-molecule of substance in a quantity of solvent so large that any further dilution would cause no thermal effect. For example, the heat of solution of sulphuric acid (H_2SO_4) in water is +75 000 J (the plus sign denoting that heat is evolved); the heat of solution of ammonium chloride (NH_4Cl) is –16 500 J (the minus sign shows that heat is absorbed).

The mutual solubility of two substances usually has definite limits: no more than a certain amount of solute can dissolve in a given quantity of solvent. A solution containing the maximum possible quantity of solute is said to be *saturated*. If further

solute is added to such a solution, it will not dissolve, and we can therefore say that a saturated solution is one which is in thermal equilibrium with the pure solute.

The concentration of the saturated solution is a measure of the ability of a given substance to dissolve in the solvent concerned, and is also called simply the *solubility* of the substance.

The solubility in general depends on the temperature. By means of Le Chatelier's principle we may relate the nature of this dependence to the sign of the heat of solution. Let us suppose that dissolution is accompanied by absorption of heat (as when ammonium chloride is dissolved in water), and that we have a saturated solution in equilibrium with undissolved solid. If this system is heated it will no longer be in equilibrium, and processes must occur which tend to oppose the external interaction (heating) which has brought the substance out of equilibrium. In the present case this means that the solubility of the substance in water will increase so as to allow further dissolution, accompanied by absorption of heat.

Thus, if dissolution is accompanied by absorption of heat, the solubility increases with temperature, but if heat is evolved on dissolution an increase in temperature will cause a decrease in solubility.

The dissolution of a gas in a liquid is usually accompanied by a large decrease in volume: the volume of the solution is considerably less than the sum of the original volumes of the solvent and the dissolved gas (for example, when one mole of nitrogen is dissolved in a large quantity of water at room temperature and atmospheric pressure the volume of the liquid increases by only 40 cm^3 , whereas the volume of this amount of gas is $22\,400 \text{ cm}^3$). Hence it follows, by Le Chatelier's principle, that the solubility of a gas in a liquid increases with the gas pressure over the liquid, at a given temperature.

The way in which the solubility of a gas depends on its pressure is easily established for weak solutions. [*Weak* (or *dilute*) solutions are those in which the number of solute molecules is small in comparison with the number of solvent molecules.] For this purpose we use the fact that thermal equilibrium (in this case, equilibrium between the gas and its saturated solution) is a dynamic equilibrium on a molecular scale. This means that, after equilibrium has been reached, the gas molecules pass from

the gas to the solution and back, but the number of molecules entering the solution from the gas per unit time is equal, in equilibrium, to the number of gas molecules leaving the solution per unit time. The number of gas molecules entering the liquid is proportional to the number of collisions per unit time between gas molecules and the liquid surface. This number in turn is proportional to the density of the gas (at a given temperature), and therefore to its pressure. Similarly, the number of gas molecules leaving the solution is proportional to its concentration. Thus, from the equality of the two numbers, it follows that the concentration of a saturated solution, i.e. the solubility of the gas, is proportional to the gas pressure over the solution (*Henry's law*).

It should be mentioned that this law is valid only for a weak solution, since in other solutions the foregoing arguments are invalid on account of the interaction between gas molecules in the solution. In consequence of this interaction the number of these molecules leaving the solution can no longer be assumed to be simply proportional to the concentration. Henry's law is therefore applicable, for instance, to oxygen and nitrogen, whose solubility in water is low, but not to the dissolution of carbon dioxide or ammonia, which are readily soluble in water.

In the great majority of cases, the dissolution of a gas is accompanied by evolution of heat; this is a quite natural result of the passage of molecules from a gaseous medium, where the interaction between molecules is weak, to a medium where the gas molecules are subject to a strong attraction exerted by solvent molecules. For this reason the solubility of gases in liquids decreases with increasing temperature (at a given pressure).

§78. Mixtures of liquids

Substances which are so rarefied that the interaction between their molecules is unimportant can mix freely with one another. In this sense we may say that all gases mix in any proportions.

In the mixing of liquids, however, various cases can occur. There exist liquids which mix in any proportions, for example alcohol and water, but the mutual solubility of other liquids is limited to various extents. For instance, water and paraffin are almost insoluble in each other; not more than 8% (by weight) of ether can be dissolved in water at room temperature, and so on.

The mutual solubility properties of liquids can be conveniently represented graphically by plotting as abscissa the concentration c of the mixture (e.g. in percent by weight) and as ordinate the temperature (or the pressure, if we are considering the dependence of solubility on pressure at a given temperature).

Figure 97 shows a diagram of this kind for a mixture of water and phenol (C_6H_5OH). One of the vertical axes corresponds to 0% water, i.e. pure phenol, and the other to pure water.

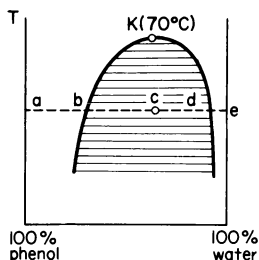


FIG. 97.

All points outside the hatched region of the diagram correspond to homogeneous mixtures of the two components; the curve forming the boundary of the hatched region represents the limit of their miscibility. For example, at the temperature corresponding to the horizontal line ae , the point b gives the limiting solubility of water in phenol, and the point d that of phenol in water. If water and phenol are mixed in quantities corresponding to a point c within the hatched region, the liquid separates into two horizontal layers with the denser layer below and the less dense one above. These two liquid layers which coexist in equilibrium represent two different phases. One is a saturated solution of water in phenol (represented by the point b), and the other is a saturated solution of phenol in water (the point d). It is easily shown, in exactly the same way as in §66 for the liquid-vapour phase diagram, that the quantities of the two phases will again be determined by the lever rule: they are inversely proportional to the lengths of cb and cd .

If the mutual solubility of two liquids increases with temperature, a point may be reached at which their miscibility becomes unrestricted. This occurs, for example, with phenol and water: at temperatures above $70^\circ C$, the two liquids mix in any proportions.

This limiting temperature is called the *critical temperature* of mixing, and the corresponding point *K* in the phase diagram (Fig. 97) is called the *critical point* of mixing; the properties of this point are in many ways similar to the critical point in the equilibrium between a liquid and a gas.

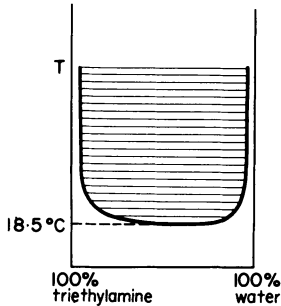


FIG. 98.

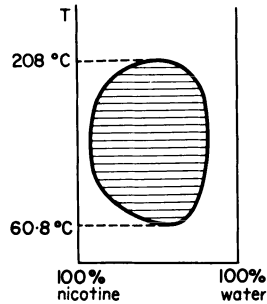


FIG. 99.

There are also cases where the critical point is not the upper but the lower limit of the region of restricted miscibility of two liquids, for example water and triethylamine ($\text{N}(\text{C}_2\text{H}_5)_3$), which mix in any proportions at temperatures below a certain critical temperature (Fig. 98). Finally, in some cases there are two critical temperatures, an upper and a lower, between which the mutual solubility of the two liquids is restricted. This occurs, for instance, with water and nicotine (Fig. 99).

§79. Solid solutions

Some substances are capable of forming crystals containing the atoms of two different substances. These are called *solid solutions* or *mixed crystals*. The ability to form solid solutions is especially common among metals, which form alloys with one another.

Mixed crystals may be referred to as solid solutions because the composition of the crystals can vary over a more or less wide range, whereas for crystals which are "chemical compounds" the composition must be entirely definite. The crystal structure of a solid solution is directly related to that of one or other component, but a chemical compound has a structure of its own.

The great majority of solid solutions are of what is called the *substitution type*. Such a solution is obtained by replacing some

of the atoms in the crystal lattice of one substance by atoms of the other substance. For such a replacement to be possible, the atoms of the new substance must of course be of about the same size as those of the solvent. Substitution-type solutions include, in particular, the majority of metal alloys. There are even cases of unrestricted mutual solubility of the two components of an alloy (for example, alloys of copper and gold); for this to be so it is evidently necessary that the two components should have crystal lattices of the same type.

Solid solutions of the substitution type can be formed not only by elements but also by chemical compounds, in which case the phenomenon is called *isomorphism*. In such mixed crystals the atoms of one compound are replaced by atoms belonging to the other compound.

It is not necessary that the two compounds should be chemically similar in order to form solid solutions. The molecular structure of the two substances must be of the same type, however. Thus, as well as chemically similar isomorphous substances (such as ZnSO_4 and MgSO_4), we also find pairs of isomorphous substances which are chemically not at all similar: BaSO_4 and KMnO_4 , PbS and NaBr , etc.

For isomorphism, not only must the molecular structure be of the same type, but the crystal lattices must also be of the same type and have similar dimensions. For example, the significance of the dimensions is seen from the compounds KCl , KBr and KI , all of which have lattices of the same type (NaCl type) but with different distances between adjoining atoms (3·14, 3·29 and 3·52 Å respectively). The comparatively small difference between the KCl and KBr lattices enables these compounds to form solid solutions of any composition, but the larger difference between KCl and KI has the result that their mutual solubility is restricted. An even greater difference may entirely prevent isomorphism.

Another type of solid solution is the *interstitial type*. In these crystals the solute atoms penetrate between the solvent atoms, slightly increasing the distance between them. In other words, they occupy positions in the lattice which are not occupied in the pure solvent. Such solid solutions can, of course, exist only if the atoms of the solute are considerably smaller than those of the solvent.

Solid solutions of the interstitial type are formed, for example, by hydrogen, nitrogen and carbon in certain metals. For instance, carbon can dissolve (at high temperatures) in what is called γ iron, a modification of iron with a face-centred cubic structure; in the resulting solution, called austenite, the carbon atoms occupy positions at the midpoints of the edges of the cubic cells, between the iron atoms at the vertices and face centres of these cells. Up to about 10% of such positions can be occupied.

§80. Osmotic pressure

If two solutions of different concentrations are separated by a porous partition, then both the solvent and the solute will pass through the partition until the two solutions are completely mixed. There are other partitions, however, which have selective transmission, i.e. allow some substances to pass through but not others; they are said to be *semipermeable*. These include various animal and vegetable membranes, colloidal films, and partitions of porous clay or porcelain in which the pores are closed by films of copper ferrocyanide ($\text{Cu}_2\text{Fe}(\text{CN})_6$). All these transmit water but retain substances dissolved in it. The passage of the solvent through such a partition is called *osmosis*.

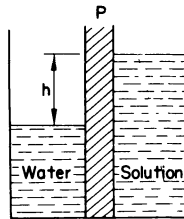


FIG. 100.

If two vessels are separated by a semipermeable partition (P in Fig. 100), and one vessel contains a solution of sugar, say, in water, and the other contains pure water, then water is found to enter the vessel containing the solution; the solution as it were attracts the solvent. This will continue until a certain difference in level is established between the water and the solution. The pressures in the two vessels are then unequal: in the vessel containing solution there is an excess pressure equal to the hydrostatic pressure of the extra column of liquid in that vessel. This excess pressure is called the *osmotic pressure* of the solution.

The reason for this phenomenon is easily understood. Since only water can pass through the semipermeable partition, equilibrium of the liquids in the two vessels does not require equality of the total pressures on the two sides of the partition. Equilibrium is reached, roughly speaking, when the pressure in the vessel containing pure water becomes equal to that part of the pressure of the solution which is due to the water molecules. The total pressure in the solution will then exceed that in the other vessel by an amount which may be regarded as the pressure due to the sugar molecules. This is the osmotic pressure of the solution.

If the solution is weak, the molecules of solvent are in general far apart, and therefore interact only very weakly with one another (though they interact, of course, with the solvent molecules). In this respect the solute molecules in a weak solution may be said to behave similarly to the molecules of an ideal gas. This in turn leads to a number of analogies between the properties of weak solutions and those of ideal gases.

We know that the pressure of an ideal gas is given by the formula $p = NkT/V$. It is found that the osmotic pressure p_{osm} of a weak solution is given by an analogous formula,

$$p_{\text{osm}} = nkT/V,$$

where V is the volume of the solution and n the number of molecules of solute in it. This is van't Hoff's formula.

It should be emphasised that the osmotic pressure of a weak solution (for a given volume and temperature) is determined only by the number of the solute particles and does not depend on their nature (or on the nature of the solvent), just as the pressure of an ideal gas is independent of its nature. As an example, we may mention that the osmotic pressure of a solution with a concentration of 1/10 mole/litre is 2·24 atm. The osmotic pressure of sea water is about 2·7 atm.

If we have a weak solution of several substances in the same solvent, then, from the above discussion, the osmotic pressure of the solution is determined by the total number of dissolved particles. It is therefore equal to the sum of the "partial" osmotic pressures of the individual solutes (corresponding to Dalton's law for gases). This should be borne in mind also when the

dissolution is accompanied by decomposition of the molecules into parts (dissociation); this phenomenon will be discussed in §§89 and 90. The osmotic pressure of such a solution depends not only on the total quantity of solute but also on the degree to which its molecules dissociate.

The analogy between a weak solution and an ideal gas extends further. For example, the height distribution of solute molecules in a field of gravity is given by a formula similar to the barometric formula (§54). This effect may be observed particularly clearly by using, instead of an ordinary solution, an emulsion consisting of very small particles of a substance suspended in a liquid. Since the mass of such particles is many times greater than that of the individual molecules, the variation of their concentration with height is seen from the barometric formula to be much more rapid, and is therefore easily observed directly. [In the barometric formula we must, of course, substitute the mass m of a particle in the emulsion minus the mass m_0 of the liquid displaced by it, in accordance with Archimedes' principle.]

§81. Raoult's law

We know that, for a given pressure, there is a definite temperature, the boiling point, at which a liquid changes into a vapour. Let us now suppose that a non-volatile substance (i.e. one which does not vaporise when a solution of it in the liquid evaporates) is dissolved in the liquid; for example, sugar dissolved in water. It is found that the boiling point of the solution is different from that of the pure solvent (at the same pressure).

From Le Chatelier's principle it is easy to deduce that the boiling point is raised when a solute is added. Let us consider a solution of sugar in water, in equilibrium with the vapour, and let a further quantity of sugar be added to the solution. The concentration of the solution is increased, and the system is no longer in equilibrium. Processes must occur in it which tend to oppose the external interaction, i.e. to decrease the concentration. For this to be so, the boiling point must rise, so that some of the vapour condenses into water.

The rise in the boiling point of the solution is shown in the pT diagram by the fact that the evaporation curve of the solution (curve 2 in Fig. 101) is somewhat to the right of curve 1, the evaporation curve of the pure solvent. At the same time, as the

diagram shows, curve 2 lies below curve 1. This means that the saturated vapour pressure of the solvent above the solution is less than that of the pure solvent at the same temperature. The decrease δp in the saturated vapour pressure and the rise δT in the boiling point when the solute is present are shown in the diagram by the vertical and horizontal lines between the two curves. These quantities can be calculated if the solution is a weak one, as will be assumed below.

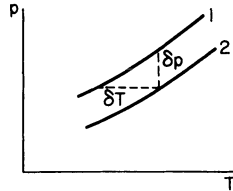


FIG. 101.

Let us return to the equilibrium shown in Fig. 100 between pure water and a solution, separated by a semipermeable partition, and suppose that the whole system is in a closed space filled with saturated water vapour. Since the gas pressure in a field of gravity decreases with increasing height, the vapour pressure over the surface of the water will be greater than that over the solution, in accordance with the foregoing discussion. The pressure difference δp is clearly that due to a column of vapour of height h :

$$\delta p = \rho_v g h,$$

where ρ_v is the density of the vapour. The height h is determined by the osmotic pressure p_{osm} of the solution: the pressure of the column of liquid balances the pressure p_{osm} . The formula for the osmotic pressure gives

$$\rho_l g h = p_{\text{osm}} = nkT/V_l;$$

we shall take n to be here the number of molecules of solute per unit mass of liquid, so that V_l is the specific volume of the liquid, $V_l = 1/\rho_l$. Hence, substituting $gh = nkT$ in the expression for δp , we obtain

$$\delta p = \rho_v nkT = nkT/V_v.$$

Finally, regarding the vapour as an ideal gas, we have its specific volume $V_v = NkT/p$, where N is the number of molecules per unit mass of vapour or, what is the same thing, per unit mass of water. The final result is

$$\delta p/p = n/N.$$

This is the required formula: the relative decrease in the vapour pressure is equal to the molecular concentration of the solution, i.e. the ratio of the numbers of solute and solvent molecules (or, what is the same thing, the ratio of the numbers of gram-molecules). This is called *Raoult's law*. We see that the change in the vapour pressure over the solution is independent of the specific properties of the solvent and solute; only the numbers of molecules are involved.

This latter property does not hold good for the other quantity under discussion, the rise δT in the boiling point. This is easily found by noting that the small quantities δp and δT are related by

$$\delta p = (dp/dT)\delta T.$$

Using the Clausius–Clapeyron equation

$$dp/dT = q/RT^2$$

(where q is the molar heat of evaporation; see §68), we obtain

$$\delta T = (RT^2/qp)\delta p.$$

Finally, substituting $\delta p/p = n/N$, we find

$$\delta T = RT^2n/qN.$$

The presence of a solute also affects the freezing point of a liquid. In the great majority of cases the solute does not enter the solid phase, i.e. pure solvent freezes out of the solution. Just as for evaporation, we can use Le Chatelier's principle and easily prove that the presence of solute lowers the freezing point. It is also found that the quantitative formula for the amount of this depression δT is the same as the formula derived above for the

change in the boiling point, q now signifying the molar heat of fusion of the solvent.

The lowering of the freezing point is often used to determine the the molecular weight (the *cryoscopic* method). After dissolving a known weight of the substance under investigation, we determine δT , and hence calculate from the above formula the number of dissolved molecules, and so the molecular weight. The molecular weight can similarly be determined from the rise in the boiling point.

§82. Boiling of a mixture of liquids

When a mixture of two liquids boils, both components of the mixture generally vaporise, so that we have an equilibrium of a liquid and a gaseous phase each of which is a mixture. The resulting phenomena can be most clearly represented by means of a phase diagram, with the concentration c of the mixture on one axis and the temperature T or the pressure p on the other. Here we shall consider cT diagrams for a given value of the pressure.

There are various types of phase diagram for the boiling of a liquid mixture. Here we shall consider those which occur for substances that mix in any proportions in the liquid state.

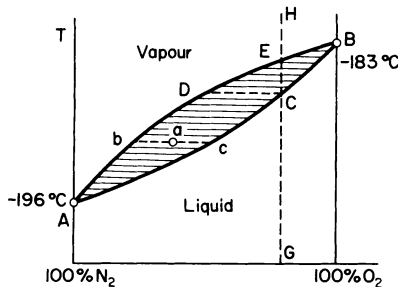


FIG. 102.

As an example of the first type we take a mixture of liquid oxygen and liquid nitrogen (Fig. 102). One of the vertical axes in the diagram corresponds to pure nitrogen and the other to pure oxygen, and between them lie all intermediate concentrations.

The region above the upper curve corresponds to states of the high-temperature phase, i.e. the gaseous mixture, and the

region below the lower curve corresponds to states of the liquid mixture. The hatched region between the two curves corresponds to equilibrium between liquid and vapour, the conditions of the liquid and vapour in equilibrium being determined by the points of intersection of the horizontal line through a given point with the two curves. For example, at the point a equilibrium exists between a gas whose composition is given by the abscissa of the point b and a liquid represented by the point c ; the relative quantities of gas and liquid are inversely proportional to the lengths of ab and ac . The upper curve ADB is called the *vapour curve* and the lower curve ACB the *liquid curve*. A phase diagram of this shape is often referred to as a "cigar".

The points A and B represent the boiling points of pure nitrogen and oxygen. Suppose that we have a liquid mixture whose composition corresponds to the vertical line GH in Fig. 102. When such a mixture is heated, its state will vary along the line GC until the point C is reached. At this temperature the liquid begins to boil, but the composition of the vapour which boils off is not the same as that of the liquid: it is the composition which can be in equilibrium with the liquid at this temperature, i.e. that which is given by the point D . Thus the vapour which boils off has a higher nitrogen concentration than the liquid. Accordingly, the composition of the liquid will move towards an increasing content of oxygen. On further heating, therefore, the point representing the state of the liquid will move upwards along the curve CB . The vapour which boils off will be represented by a point which moves upwards along the curve DB .

We see that the mixture does not boil at a constant temperature, unlike a pure liquid. The point at which boiling ceases depends on the conditions under which it occurs. If the vapour which boils off remains in contact with the liquid, the total composition of the liquid and vapour remains fixed, and the states of the system are always represented by points on the line GH . Hence we see that boiling begins at the point C and ends at the temperature of the point E where the vertical line GH intersects the upper curve of the "cigar".

If, however, boiling takes place in an open vessel and the vapour which boils off is steadily removed, only the vapour which has just boiled is in equilibrium with the liquid at any instant. A quantity of liquid, on boiling, is converted entirely

into vapour, i.e. the resulting quantity of vapour must have the same composition as the boiling liquid. Thus the last part of the boiling process occurs at a point where the composition of the liquid and that of the vapour are the same, i.e. at the boiling point *B* of pure oxygen.

Exactly similar effects occur in the condensation of a vapour into a liquid.

Another type of phase diagram occurs, for example, for a mixture of chloroform and acetone (Fig. 103). This differs from the previous case in that the two curves have a maximum point *A*, at which they touch. Here again, the region between the curves corresponds to liquid and vapour in equilibrium, while the regions above and below the curves correspond to the gaseous and liquid phases.

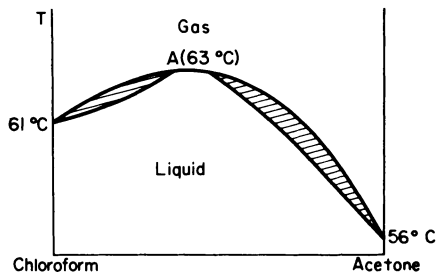


FIG. 103.

Boiling or condensation occurs in a similar manner to the preceding case. For example, when liquid boils in an open vessel, the points representing the states of the liquid and the vapour move upwards along the two curves, but the process now terminates not at the boiling point of one of the pure components but at the point *A* where the curves touch. At this point the composition of the liquid is the same as that of its vapour. A mixture whose composition corresponds to the point *A* (called an *azeotropic* mixture) therefore boils away completely at a constant temperature, as if it were a pure substance.

Finally, there are mixtures (for example, carbon disulphide and acetone) whose phase diagrams differ from the preceding type only in that the curves have minima instead of maxima (Fig. 104).

The effects described above are widely used in practice in order to separate the components of various mixtures. In its

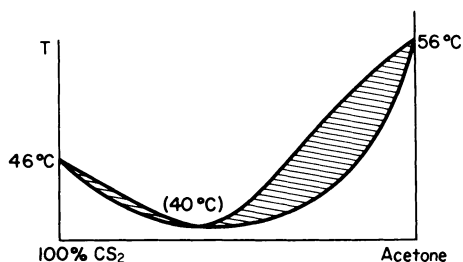


FIG. 104.

simplest form, the method of *fractional distillation* consists in collecting and condensing the initial fractions of the vapour boiling off from a liquid mixture, and then redistilling the resulting substance. For example, when a mixture of alcohol and water boils, the vapour formed has a higher content of the more volatile alcohol than is present in the liquid. By condensing the initial fractions of this vapour and again boiling the resulting liquid, we can separate the water and the alcohol more and more completely. When the phase diagram is of the type shown in Fig. 102 the components of the mixture can, in principle, be completely separated by repeating the process several times. When the phase diagram is as in Fig. 103 or 104, however, complete separation is not possible: only an azeotropic mixture can be separated, together with one or the other pure substance, depending on the composition of the original mixture. The mixture of water and alcohol mentioned above is of this type; it has a minimum boiling point at a composition of 95.6% by weight of alcohol. The alcohol cannot be further purified by fractional distillation.

§83. Reverse condensation

The existence of critical points for liquid-gas transitions in pure substances has the result that critical phenomena occur in mixtures also. Without analysing all possible variations, let us consider some characteristic features of these phenomena.

The phase diagram of oxygen-nitrogen mixtures shown in Fig. 102 refers to a pressure of 1 atm. At higher pressures the diagram remains of the same type, but only up to the critical pressure of one of the pure components, in this case nitrogen, at 33.5 atm (the critical pressure of oxygen is 49.7 atm). Since

pure nitrogen cannot separate into phases above this point, it is evident that the “cigar” in the phase diagram of the mixtures must become “detached” from the vertical axis and become of the form shown in Fig. 105. We see that a point K is now present in the diagram at which the two coexisting phases become identical; this is called the critical point. Here again the presence of the critical point means that a continuous transition is possible between liquid and gas, so that the distinction between these two phases becomes arbitrary.

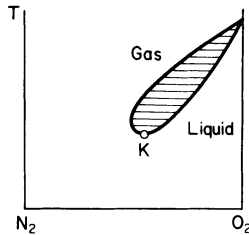


FIG. 105.

The condensation of a gas mixture may be accompanied by unusual phenomena when there is a critical point in the phase diagram. We may illustrate these by means of a cp diagram (for a given value of the temperature), which more nearly corresponds to the usual conditions under which they are actually observed.

Figure 106 shows part of such a phase diagram near the critical point K ; unlike the cT diagrams used above, the gas phase corresponds to the region *below* the hatched area, i.e. the region of low pressures.

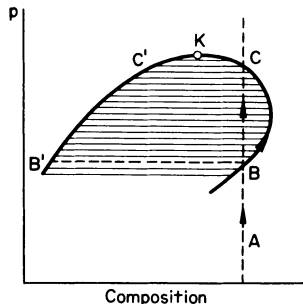


FIG. 106.

Let us consider a mixture whose composition corresponds to the vertical line AC . In isothermal compression, when the point B is reached, the mixture begins to condense, forming a liquid phase B' . As the pressure is further increased, the quantity of liquid at first increases but later decreases, and when the point C is reached the liquid (which is then represented by the point C') disappears entirely. This phenomenon is called *reverse condensation*.

§84. Solidification of a mixture of liquids

The phase diagram for a liquid and a solid can be represented in the same way as that for a liquid and a gas. We again plot the concentration of the mixture (percent by atoms) as abscissa, and the temperature as ordinate, and consider the diagram for a given pressure.

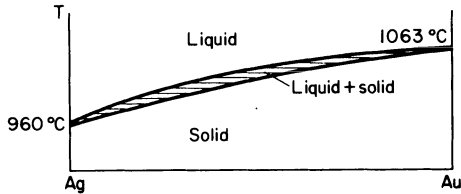


FIG. 107.

If the two substances mix in any proportions both in the liquid and in the solid state, then the form of the diagrams is exactly similar to the liquid-gas phase diagrams discussed in §82. For example, an alloy of silver and gold has the phase diagram shown in Fig. 107. The region above the curves corresponds to liquid mixtures of the two metals, and the region below the curves corresponds to solid alloys. The process of melting of the alloy occurs similarly to the boiling of a liquid mixture as described in connection with Fig. 102.

The phase diagram for the bismuth-cadmium system, shown in Fig. 108, is of an entirely different type. This system has the property that the two components form no solid solutions.

The region which is not hatched corresponds to liquid mixtures. In all the other regions there is separation into two phases. In region I the two phases are solid crystals of pure cadmium (represented by the left-hand vertical axis) and a liquid (re-

presented by the curve AO). For example, at a point d in this region there is equilibrium between phases represented by the points where the horizontal line ef meets the ordinate axis (pure cadmium) and the curve AO (liquid mixture); the quantities of these phases are inversely proportional to the lengths de and df . Similarly, in region II the solid phase is bismuth in equilibrium with a liquid whose composition is determined by the curve OB . Finally, in region III there is a mixture of solid crystals of cadmium and bismuth.

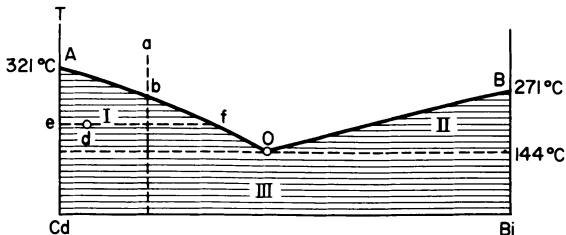


FIG. 108.

The points A and B are the melting points of pure cadmium and pure bismuth. The curve AOB gives the temperature at which liquid mixtures of the two components begin to solidify.

Let us consider, for instance, the process of solidification of a liquid mixture whose composition is given by the vertical line ab . The solidification begins at the temperature of the point b at which this vertical line intersects the curve AO , and crystals of cadmium separate from the liquid. As the temperature decreases further, the liquid mixture becomes richer in bismuth, and the point representing it moves downwards along the curve bO until the point O is reached. The temperature then remains constant until the liquid has completely solidified. At the temperature of O , crystals of the remaining cadmium and all the bismuth that was in the liquid are formed.

The point O is called the *eutectic point*. It is a point at which three phases are in equilibrium: solid cadmium, solid bismuth, and the liquid mixture. The crystalline mixture which solidifies out at the eutectic point consists of very small crystals of each component, called a eutectic mixture. To the right of O , in region III, the eutectic mixture contains larger crystals of previously solidified bismuth, and to the left, crystals of cadmium.

Figure 109 shows the typical form of the “cooling curve” corresponding to Fig. 108. Here the temperature of the system is plotted as a function of time during slow cooling of a liquid of given composition (in this case corresponding to the vertical line ab). When the point b is reached, a break occurs on the cooling curve; because solidification begins, which is accompanied by the evolution of heat, the cooling becomes somewhat slower. At the temperature of the eutectic point there is a “plateau”: the curve has a horizontal section, corresponding to the complete solidification of the alloy at a constant temperature. The recording of such cooling curves is the basis of the derivation of phase diagrams by the method of *thermal analysis*.

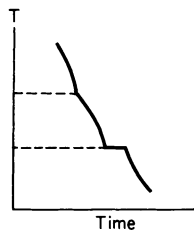


FIG. 109.

The phase diagram of the silver-copper system, shown in Fig. 110, differs from the previous one in that each of its solid components can dissolve a certain amount of the other. The diagram therefore contains three single-phase regions: in addition to region I (liquid mixtures), we have region II (solid solutions of copper in silver) and region III (solid solutions of silver in copper).

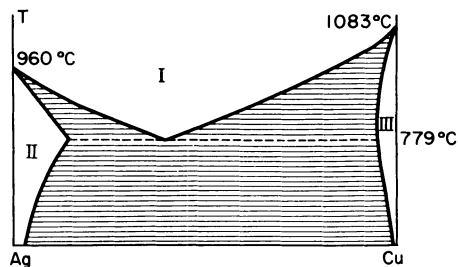


FIG. 110.

Finally, let us consider the phase diagram of the aluminium-calcium system (Fig. 111). In this case, although the two components do not form solid solutions, there exist certain chemical compounds; we might say that only solid solutions of certain definite compositions exist. The vertical line BD corresponds to the compound CaAl_2 . The point B is the melting point of this compound, and is the maximum of the curve ABC . Another compound, CaAl_4 , decomposes before melting begins. The vertical line EF corresponding to this compound therefore does not reach the boundary AB of the liquid state. All the hatched areas are regions of separation into two phases. The two phases that are in equilibrium are always given by the points of intersection of a horizontal line with the two nearest vertical lines in the diagram. For example, in region I the liquid is in equilibrium with crystals of the compound CaAl_2 ; in region II, with crystals of CaAl_4 ; in region III, crystals of aluminium with crystals of CaAl_4 , and so on.

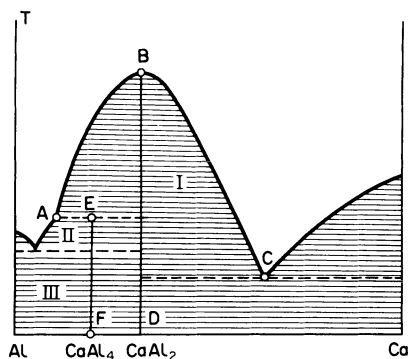


FIG. 111.

We may note that the study of phase diagrams by means of thermal analysis can itself provide information as to the existence of solid chemical compounds of various substances. The existence of a compound is shown by the presence on the melting curve of a maximum (as at B in Fig. 111) or a break (as at A).

There is a great variety of phase diagrams for different mixtures. The few diagrams described here are among the simplest. These examples show, however, the characteristic properties and types which may be seen in more complicated diagrams also.

§85. The phase rule

We shall now recapitulate and generalise some of the properties of phase equilibrium described in Chapters IX and X.

The thermal state of a homogeneous body consisting of a single substance is defined by the values of two independent quantities, the temperature T and the pressure p . If a further phase of the same substance is added (e.g. ice and water), these phases can coexist not at all values of p and T but only when a certain relation holds (represented by a curve in the pT diagram). We may say that the equilibrium with ice imposes on the equation of state of water a further condition, as a result of which the number of independent quantities is reduced from two (p and T) to one (p or T).

Three phases of one substance (e.g. water, ice and steam) can coexist only for certain definite values of p and T , where the ice–water and steam–water equilibrium curves intersect. We may say that the addition of a third phase imposes a further condition, as a result of which the number of independent quantities is reduced to zero.

Hence it is clear that four phases of one substance (e.g. water, steam, and two forms of ice) can not exist in equilibrium. Such an equilibrium would require three added conditions to be satisfied, and this can not be achieved by means of the two disposable variables p and T .

Let us now consider a body consisting of two substances, such as a liquid solution. Its state is defined by three independent variables: the temperature T , the pressure p and the concentration c . Let this solution be in equilibrium with its vapour (containing the same two substances). This imposes a certain further condition, and only two of the three quantities describing the state of the solution remain arbitrary. Thus equilibrium between a liquid solution and a vapour is possible at any pressure and temperature (for example), but the concentration of the solution, and therefore that of the vapour, must then have a definite value. We have already seen this from the phase diagrams in the present chapter.

If a further phase is added which consists of the same two substances, this imposes a further condition, and only one variable remains arbitrary. For instance, at a given pressure three phases can coexist at only one point, with a definite tem-

perature and definite concentrations. The eutectic point in the phase diagrams in §84 is a point of this type.

Finally, four phases of two components can be in equilibrium only for certain values of all the quantities (pressure, temperature and concentrations), and the equilibrium of five (or more) phases is impossible.

These statements are easily generalised to the equilibrium of phases containing any number of components. Let the number of components be n , and the number of coexisting phases r ; and let us consider one phase. Its composition is specified by the values of $n - 1$ concentrations, for example the ratios of the quantities of each of $n - 1$ components to that of the n th. Thus the state of the phase is defined by $n + 1$ quantities: p , T and $n - 1$ concentrations. But this phase is in equilibrium with $r - 1$ other phases, which imposes $r - 1$ extra conditions on its equation of state. These must not exceed the number of variables, i.e. $n + 1$ must not be less than $r - 1$, or in other words

$$r \leq n + 2.$$

Thus not more than $n + 2$ phases composed of n substances can coexist in equilibrium. This is called the *phase rule*.

When the maximum number ($n + 2$) of phases coexist, all the quantities which describe the states of the phases (p , T and the concentrations of all the phases) must have definite values. When r phases are in equilibrium, the values of $(n + 1) - (r - 1) = n + 2 - r$ quantities may be specified arbitrarily.

CHAPTER XI
CHEMICAL REACTIONS

§86. **Heats of reaction**

This chapter will deal with chemical reactions from the physical point of view, having regard to properties which are common to all reactions, whatever the chemical nature of the reacting substances.

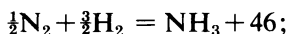
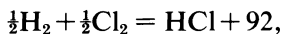
Any chemical reaction is accompanied by absorption or evolution of heat. In the former case the reaction is said to be *endothermic*, and in the latter case *exothermic*. It is clear that, if a reaction is exothermic, the reverse reaction is endothermic, and *vice versa*.

The heat involved in a reaction depends in general on the conditions under which it occurs. Hence, strictly speaking, we should distinguish the quantity of heat according as the reaction occurs at constant pressure or at constant volume. In practice, however, the difference is usually very slight.

The heat of reaction is written in the reaction equation with a positive sign on the side where heat is evolved, or with a negative sign where it is absorbed. For example, the equation



signifies that 400 kJ of heat are evolved in the combustion of one gram-atom of carbon (graphite). Two further examples are



here, as in all subsequent examples, the heat is again stated in kilojoules per mole of the reacting substances.

In the above examples it has been assumed that all the substances (except graphite) are in the gaseous state (at room

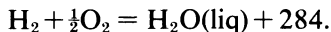
temperature and atmospheric pressure). The aggregate state of the reacting substances must be specified, since the heat of reaction depends on this state, and the dependence may be very considerable. As an example, let us find the difference between the heats of formation of liquid water and steam from gaseous oxygen and hydrogen. The heat of evaporation of a gram-molecule of water (at 20°C) is 44 kJ, i.e.



Adding this to the equation of formation of steam,



we obtain the equation of formation of liquid water,

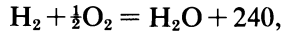
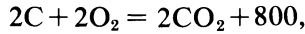


The heat of reaction also depends, of course, on the temperature at which the reaction takes place. The value is easily converted from one temperature to another if the specific heats of all the reacting substances are known, in the same way as we have just converted the value from one aggregate state to another. To make the conversion it is necessary to calculate the heat required to bring all the substances which participate in the reaction from one temperature to another.

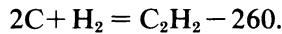
If several reactions occur in succession, it follows from the law of conservation of energy that the total heat of the whole sequence of reactions is equal to the sum of the heats of each successive reaction. We can say, moreover, that, if we start from certain given substances and produce other substances by a series of intermediate reactions, the total heat evolved is independent of the intermediate stages through which the reaction took place.

By means of this rule we can, in particular, calculate the heats of reactions which in practice could never occur. Let us find, for example, the heat of formation of acetylene gas directly from the elements carbon (graphite) and hydrogen: $2\text{C} + \text{H}_2 = \text{C}_2\text{H}_2$. This reaction cannot occur directly and in practice is brought about by other means; the heat of reaction therefore cannot be

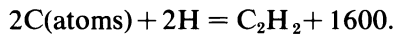
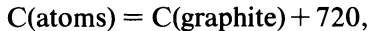
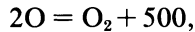
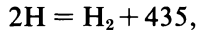
measured directly. We can, however, calculate this heat from the known (directly measurable) heats of combustion of carbon, hydrogen and acetylene itself:



Adding the first two equations and subtracting the third, we obtain



The heat of formation of a compound from its elements depends on their state. In physics the heat of formation from the atoms of the elements is of interest, not that from the elements in their natural state. The heat of formation from the atoms determines the internal energy of the compound as such and is independent of the state of the initial substances. Some examples are



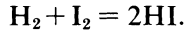
The heat of formation of a compound from the elements may be either positive or negative, but the heat of formation from the atoms is always positive, since otherwise the compound would be unstable and could not exist.

§87. Chemical equilibrium

As a chemical reaction proceeds, the quantities of the original substances decrease, and reaction products accumulate. Ultimately, the reaction leads to a state in which the quantities of all the substances no longer vary. This is called a state of *chemical equilibrium*, a particular case of thermal equilibrium.

In chemical equilibrium there is generally present a certain quantity of the original substances as well as the products formed in the reaction. It is true that in many cases this quantity is very small, but this does not, of course, affect the principle.

The establishment of chemical equilibrium in which both initial and final substances are present occurs for the following reason. Let us consider, for example, a reaction between hydrogen and iodine gases to form hydrogen iodide:



As well as the formation of HI from H_2 and I_2 , in a mixture of these three substances the reverse process of dissociation of HI into hydrogen and iodine will also necessarily occur: the forward reaction is always accompanied by the reverse reaction. As the quantity of HI increases and that of H_2 and I_2 decreases, the forward reaction will obviously become slower and the reverse reaction quicker, and a point is finally reached at which the rates of the two reactions become equal, with the same number of new HI molecules formed as dissociate in the same time. The quantities of all the substances thereafter remain unchanged.

Thus chemical equilibrium (and in fact other types of thermal equilibrium) is dynamic on the molecular scale; the reactions do not actually cease, but the forward and reverse reactions occur at equal rates and therefore produce no overall effect.

It is clear that, if the reaction in the above example begins from a mixture of hydrogen and iodine, the relative quantities of all three substances in the equilibrium state will be the same as in a reaction which begins with the decomposition of pure HI. The chemical equilibrium position does not depend on the side from which it is approached.

Moreover, the chemical equilibrium also does not depend on the conditions under which the reaction occurs or on the intermediate stages through which it passes. The position of equilibrium depends only on the state of the substances in equilibrium, i.e. on the temperature and pressure of the equilibrium mixture.

When the temperature changes, the position of chemical equilibrium is altered. The direction of this shift depends entirely on the heat of reaction, as is easily seen by means of Le Chatelier's

principle. Let us consider an exothermic reaction, such as the formation of ammonia from nitrogen and hydrogen ($\text{N}_2 + 3\text{H}_2 = 2\text{NH}_3$), and assume that the reaction has already reached a state of equilibrium. If the equilibrium mixture is now heated, processes must occur in it which tend to cool it: a certain quantity of ammonia must decompose, and heat is thus absorbed. This means that the chemical equilibrium is shifted in the direction such that the quantity of ammonia is decreased.

Thus the "yield" of exothermic reactions decreases when the temperature is raised; in endothermic reactions, on the other hand, the yield increases with increasing temperature.

Similarly, the dependence of the equilibrium position on the pressure is related to the change in volume accompanying the reaction (at constant pressure). Increasing the pressure lowers the yield of reactions in which the volume of the reacting mixture increases, and raises that of reactions in which the volume decreases. The latter case occurs, for instance, in the reaction of formation of gaseous ammonia: since the number of NH_3 molecules formed is less than the number of reacting N_2 and H_2 molecules, the volume of the gas mixture decreases in the reaction.

§88. The law of mass action

We shall now give a quantitative formulation of the concept of chemical equilibrium. Let us first consider chemical reactions in a gas mixture, all the substances participating in a reaction being in the gaseous state.

As an example we shall again take the reaction of formation of HI. The reaction between hydrogen and iodine can occur when H_2 and I_2 molecules collide. The rate of the reaction of HI formation (i.e. the number of HI molecules formed per unit time) is therefore proportional to the number of such collisions. This number in turn is proportional to the densities of hydrogen and iodine in the mixture, i.e. the numbers of molecules of hydrogen and iodine per unit volume. The density of a gas is proportional to its pressure. Thus the rate of the reaction of HI formation is proportional to the partial pressures of these gases in the mixture, i.e. is $k_1 p_{\text{H}_2} p_{\text{I}_2}$, where the coefficient k_1 depends only on the temperature. Similarly, the rate of the reaction of HI decomposition is proportional to the number of collisions

between HI molecules, and therefore to the square of the partial pressure of HI in the mixture; let it be $k_2 p_{\text{HI}}^2$.

In equilibrium, the rates of the forward and reverse reactions are equal:

$$k_1 p_{\text{H}_2} p_{\text{I}_2} = k_2 p_{\text{HI}}^2.$$

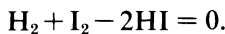
Putting $k_2/k_1 = K(T)$, we thus have

$$p_{\text{H}_2} p_{\text{I}_2} / p_{\text{HI}}^2 = K(T).$$

This equation relates the partial pressures of all three gases in the equilibrium state. The quantity $K(T)$ is called the *equilibrium constant* for the reaction concerned. It is independent of the quantities of the reacting substances. The relation expressed by the above formula is called the *law of mass action*.

This law can be written in a similar form for any other reaction between gases. It may be written in a general form as follows.

In the chemical equation for the reaction we can arbitrarily transfer all the terms to the same side, writing e.g.



Any reaction may be represented in the general form

$$\nu_1 A_1 + \nu_2 A_2 + \dots = 0,$$

where A_1, A_2, \dots are the chemical symbols of the reacting substances, and ν_1, ν_2, \dots are positive or negative integers; for instance, in the above example $\nu_{\text{H}_2} = \nu_{\text{I}_2} = 1$, $\nu_{\text{HI}} = -2$. Then the law of mass action takes the form

$$p_1^{\nu_1} p_2^{\nu_2} \dots = K(T),$$

where p_1, p_2, \dots are the partial pressures of the various gases.

It is often more convenient to use the concentration of the substances in the mixture, instead of the partial pressures. We define the concentration of the i th substance in the mixture as the ratio $c_i = N_i/N$ of the number N_i of molecules of this substance to the total number N of molecules in the mixture, or,

what is the same thing, the ratio of the corresponding numbers of moles. Since the total pressure of the gas mixture is $p = NkT/V$, where V is the volume of the mixture, and the partial pressures $p_i = N_i kT/V$, we have

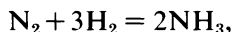
$$p_i = c_i p.$$

Substituting these expressions in the equation of the law of mass action, we obtain the latter in the form

$$c_1^{\nu_1} c_2^{\nu_2} \dots = K(T) p^{-(\nu_1 + \nu_2 + \dots)}$$

which relates the equilibrium concentrations of all the substances. The quantity on the right of this equation is also called the equilibrium constant, but may depend on the pressure as well as on the temperature. It is independent of the pressure only if the sum $\nu_1 + \nu_2 + \dots = 0$, i.e. if the total number of molecules is unchanged in the reaction (as, for example, in the reaction $\text{H}_2 + \text{I}_2 = 2\text{HI}$).

For the reaction of ammonia formation, for instance:



we have

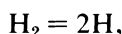
$$c_{\text{N}_2} c_{\text{H}_2}^3 / c_{\text{NH}_3}^2 = K(T) / p^2.$$

When the pressure increases, the right-hand side of this equation decreases, and the left-hand side must therefore decrease also. Thus the equilibrium concentrations of the initial substances decrease and that of ammonia increases, in agreement with the result previously found by means of Le Chatelier's principle. We have also seen that the yield of this reaction must decrease with increasing temperature; we can now say that in this case the equilibrium constant $K(T)$ increases with temperature.

The following comment must be made regarding the foregoing derivation of the law of mass action. In our discussion it has been assumed that the course of the reaction is as represented by the chemical equation. In the reaction of HI formation this is in fact so, but the majority of reactions do not really take

place in the way that would be expected from the equation; for example, the formation of an ammonia molecule does not occur by the collision of a nitrogen molecule with three hydrogen molecules. The representation of the reaction by a single equation is usually a mere summary of a sequence of intermediate steps, taking into account only the initial and final substances. We shall discuss this further below. The properties of chemical equilibrium and the law of mass action describing them are, however, independent of the actual reaction mechanism.

To illustrate the application of the law of mass action, let us make a complete analysis of the simple reaction of dissociation of hydrogen:



and determine the degree of dissociation which is reached at equilibrium. Let the total number of hydrogen atoms (both isolated and in molecules) be A . The degree of dissociation x may be defined as the ratio of the number N_{H} of hydrogen atoms to the total number of atoms A . Then

$$N_{\text{H}} = Ax, \quad N_{\text{H}_2} = \frac{1}{2}A(1-x), \quad N = N_{\text{H}} + N_{\text{H}_2} = \frac{1}{2}A(1+x).$$

Expressing the concentrations c_{H} and c_{H_2} in terms of these quantities and substituting in the law of mass action, we find

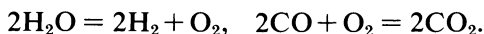
$$c_{\text{H}_2}/c_{\text{H}}^2 = (1-x^2)/4x^2 = pK,$$

whence

$$x = 1/\sqrt{1+4pK};$$

this determines, in particular, the way in which the degree of dissociation depends on the pressure.

If several different reactions can occur in a gas mixture, the law of mass action must be applied to each reaction separately. For example, in a mixture of the gases H_2 , O_2 , CO , CO_2 and H_2O , the following reactions can occur:

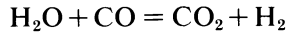


For these reactions we have

$$p_{\text{H}_2\text{O}}^2/p_{\text{H}_2}^2 p_{\text{O}_2} = K_1,$$

$$p_{\text{CO}}^2 p_{\text{O}_2}/p_{\text{CO}_2}^2 = K_2,$$

and the state of chemical equilibrium is determined by the simultaneous solution of these two equations. In the mixture considered, other reactions can occur, for example



but this reaction need not be considered, as it is just the sum of the two reactions written above, and the law of mass action would give an equation which is simply the product of the two equations previously derived.

Let us now consider a reaction which involves not only gases but also a solid. The reaction between the solid and a gas molecule can occur when the latter collides with the surface of the solid. The number of collisions of gas molecules with unit area of the surface obviously depends only on the density of the gas and not on the quantity of the solid. Accordingly, the rate of reaction per unit area of the surface of the solid will be proportional only to the partial pressure of the gases and does not depend on the quantity of the solid. Hence it is clear that the law of mass action is valid also for reactions which involve solids, with the difference that the equation for it includes only the concentrations of the gases and not the quantity of solids. The properties of the latter affect only the temperature dependence of the equilibrium constant.

For example, in the decomposition of limestone with the evolution of carbon dioxide,



the only gas is CO_2 (since the calcium oxide remains solid). The law of mass action therefore gives simply

$$p_{\text{CO}_2} = K(T).$$

This means that in equilibrium (at a given temperature) carbon dioxide over limestone must have a definite partial pressure. This is similar to the case of evaporation, where again the gas pressure over a body is determined only by the temperature and does not depend, for example, on the quantity of either substance.

The law of mass action also holds good for reactions between substances in solution if the solution is a weak one; here again we have an analogy between the properties of gases and those of weak solutions, as already noted in §80. The derivation of the law of mass action for gas reactions has been based on a calculation of the number of collisions between molecules. A similar calculation can be made for reactions in solution; the fact that the reacting molecules are not in a vacuum but in a medium, the solvent, affects only the dependence of the equilibrium constant on temperature and pressure. In the equation of the law of mass action,

$$c_1^{v_1} c_2^{v_2} \dots = K(p, T),$$

the dependence of K on both temperature and pressure therefore remains unknown. The concentrations c_1, c_2, \dots in this equation are now defined as the quantities of solutes in a given quantity (or per unit volume) of the solvent.

A similar formula is valid for reactions which involve not only the solutes but also the solvent, for example the hydrolysis of cane sugar to glucose and fructose which occurs in an aqueous sugar solution. Since the number of water molecules is much greater than the number of sugar molecules (the solution being assumed weak), the concentration of water is practically unchanged by the reaction. Thus, in the equation of the law of mass action, only the solute concentrations need be included:

$$\frac{[\text{cane sugar}]}{[\text{glucose}][\text{fructose}]} = K(T, p),$$

where the square brackets denote molar concentrations, i.e. numbers of moles per litre of water.

§89. Strong electrolytes

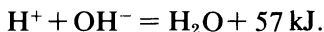
A large number of substances are present in solution not as molecules but as charged constituents of molecules called ions; the positive ions are known as *cations* and the negative ions as *anions*. Such substances are said to be *strong electrolytes*. A substance which dissolves in the form of ions is said to *dissociate* in solution, and this phenomenon is called *electrolytic dissociation*.

In solution in water, almost all salts are strong electrolytes, and so are some acids (such as HCl, HBr, HI and HNO₃) and some bases (such as NaOH and KOH). In salts the cation is the metal and the anion is the acid radical (e.g. NaCl → Na⁺ + Cl⁻). Acids dissociate to form the cation H⁺ and the acid-radical anion (HNO₃ → H⁺ + NO₃⁻); alkalis give the metal cation and the anion OH⁻, called *hydroxyl* (NaOH → Na⁺ + OH⁻).

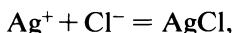
The phenomenon of electrolytic dissociation is observed in some other solvents also, but it appears most strongly in aqueous solutions.

If two strong electrolytes are simultaneously dissolved in water, for instance NaCl and KBr, there is no reason to regard the solution as one of NaCl and KBr specifically; it contains only the separate ions K⁺, Na⁺, Cl⁻ and Br⁻, and the same solution could just as well (or rather, with just as little meaning) be called a solution of NaBr and KCl.

In reactions between strong electrolytes in solution, only the separate ions actually take part, since there are no undissociated molecules in the solution. The heat of reaction between strong electrolytes therefore depends only on the ions which directly participate in the reaction, and not on which other ions are present in the solution (if the solution is weak, of course). Let us consider, for example, the neutralisation of strong acids with alkali. It is inaccurate to write the neutralisation reaction as, for example, NaOH + HCl = NaCl + H₂O; in reality, only the H⁺ and OH⁻ ions react, combining to form water: H⁺ + OH⁻ = H₂O. This reaction is obviously the same for all strong acids and alkalis, whatever the metal and the acid radical. The heat of reaction is therefore likewise the same for the neutralisation of any strong acid by any strong alkali. The value of this heat for one mole of acid and one mole of alkali is 57 kJ:



Let us next consider a saturated solution of any strong electrolyte whose solubility is low, for example silver chloride in water. By the definition of saturation this solution is in equilibrium with solid silver chloride. This equilibrium may be regarded as the chemical equilibrium of the reaction



where Ag^+ and Cl^- are in the solution and AgCl in the solid state: the number of AgCl molecules which go into solution per unit time is equal to the number of molecules deposited from the solution by combination of ions per unit time. Since, on the other hand, the solution is weak (because the solubility of silver chloride is low), we can apply the law of mass action. Here, as explained previously, only the concentrations of the solutes need be included, and we find

$$[\text{Ag}^+][\text{Cl}^-] = K,$$

where the square brackets again denote molar concentrations (numbers of moles per litre of water). The constant K (which is, of course, a function of temperature) is called the *solubility product* for the electrolyte concerned. For example, for silver chloride at room temperature $K = 1 \times 10^{-10}$ (mole/l)²; for CaCO_3 , $K = 1 \times 10^{-8}$ (mole/l)².

Thus the product of the concentrations of anions and cations in a saturated solution of a strong electrolyte of low solubility is a constant. If no salt containing silver or chlorine ions, except silver chloride itself, is present in solution in water, the concentrations $[\text{Ag}^+]$ and $[\text{Cl}^-]$ are equal to the solubility c_0 of silver chloride. Hence it follows that

$$K = c_0^2.$$

Now let a quantity of another chloride, of high solubility (such as NaCl), be added to a saturated solution of AgCl . Then some of the latter will be deposited as solid from the solution, since the addition of NaCl raises the concentration of chloride ions while that of silver ions remains unchanged; some of the AgCl must therefore be deposited in order that the product $[\text{Ag}^+][\text{Cl}^-]$ may remain constant.

§90. Weak electrolytes

As well as strong electrolytes, there are substances which dissociate in solution but do so only partly; in solutions of these substances there are not only ions but also neutral molecules. Such substances are called *weak electrolytes*. The majority of acids and bases, and some salts (such as HgCl_2), are weak electrolytes in aqueous solution.

The law of mass action is applicable to weak solutions of weak electrolytes. Let us consider, for example, a solution of acetic acid (CH_3COOH), which dissociates in water according to the equation



(the symbol Ac denoting the acid radical CH_3COO). Dissociation continues until equilibrium is established, when the ion concentrations satisfy the equation

$$[\text{Ac}^-][\text{H}^+]/[\text{HAc}] = K.$$

The constant K is called the *dissociation constant*. For instance, for acetic acid at room temperature $K = 2 \times 10^{-5}$ mole/litre.

A dissociation reaction is endothermic, i.e. it occurs with absorption of heat. As with all endothermic reactions, its "yield" increases with rising temperature, i.e. the dissociation constant increases.

The dissociation constant is independent of the quantity of dissolved electrolyte (so long as the solution remains weak) and is a fundamental property of the electrolyte, but the *degree of dissociation* (i.e. the ratio of the number of dissociated molecules to the total number of electrolyte molecules) depends on the concentration of the solution.

Let a total of c moles of electrolyte be dissolved in a litre of water, and let the degree of dissociation be α . Then the number of dissociated moles is $c\alpha$. If an electrolyte molecule dissociates into one anion and one cation (as in the example of acetic acid considered above), then the concentration of each is $c\alpha$. The concentration of undissociated molecules is $c(1 - \alpha)$. The law of mass action therefore gives

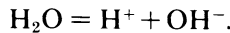
$$\alpha^2 c / (1 - \alpha) = K.$$

Hence we find the degree of dissociation in terms of the concentration of the solution:

$$\alpha = \frac{-K + \sqrt{(K^2 + 4Kc)}}{2c} = \frac{2K}{K + \sqrt{(K^2 + 4Kc)}}$$

This formula shows that, as the concentration c decreases, the degree of dissociation increases, tending to unity at infinite dilution (i.e. as $c \rightarrow 0$). Thus, the more dilute the solution, the more the electrolyte is dissociated. This naturally follows from the fact that a molecule dissociates under the action of water molecules, which are present everywhere, but for recombination to occur two different ions must come together, and this occurs more rarely in more dilute solutions.

Water is itself a very weak electrolyte. A very small fraction of its molecules are dissociated in accordance with the equation



Since H_2O is at the same time the solvent with respect to the ions H^+ and OH^- , the formula for the law of mass action need include, as we know, only the concentrations of these ions:

$$[\text{H}^+][\text{OH}^-] = K.$$

For pure water at 25°C ,

$$K = 10^{-14} \text{ (mole/litre)}^2.$$

Since in pure water the concentrations of H^+ and OH^- ions are evidently equal, we find that each is 10^{-7} . Thus one litre of water contains only 10^{-7} mole of H^+ ions (and the same quantity of OH^-); 1 mole of water (18 g) is dissociated only in ten million litres.

The decimal logarithm of the concentration of H^+ ions, with sign reversed, is called the pH:

$$\text{pH} = -\log_{10}[\text{H}^+].$$

For pure water at 25°C the pH is 7.0; at 0°C it is 7.5 and at 60°C 6.5.

When acids dissolve they release H^+ ions. But the product of concentrations $[H^+][OH^-]$ must remain constant and equal to 10^{-14} . Some of the OH^- ions must therefore combine with H^+ ions to form neutral molecules of water. Thus the concentration $[H^+]$ is greater than its value in pure water (10^{-7}), and the pH of an acid solution is consequently less than 7. Similarly, in solutions of alkalis (which release OH^- ions) the pH is greater than 7. The pH of a solution is therefore a quantitative measure of its degree of acidity or alkalinity.

Solutions containing a weak acid (such as acetic acid HAc) and a salt of it which is a strong electrolyte (e.g. sodium acetate, NaAc) have interesting properties. The completely dissociated salt yields a large quantity of Ac^- ions in the solution. From the equation of dissociation of the acid,

$$[H^+][Ac^-]/[HAc] = K,$$

we find that the presence of excess Ac^- ions in the solution causes a decrease in the number of H^+ ions, i.e. inhibits the dissociation of the acid. The concentration $[HAc]$ of undissociated acid molecules is therefore practically equal to the total concentration of the acid (denoted by c_a). The concentration of Ac^- ions, which are almost entirely supplied by the salt, is practically equal to the salt concentration (c_s). Thus $[H^+] = Kc_a/c_s$, and the pH of the solution is

$$pH = -\log_{10}[H^+] = -\log_{10}K + \log_{10}(c_s/c_a).$$

This depends only on the ratio of concentrations of the salt and the acid. Thus dilution of the solution, or the addition of small quantities of any other acids or alkalis, has practically no effect on the pH of the solution. A solution of this type whose pH remains constant is called a *buffer solution*.

§91. Activation energy

Hitherto we have considered only the state of chemical equilibrium, leaving aside the question of reaction mechanisms and rates. The calculation of the number of collisions of molecules in §88 served only to derive the conditions of equilibrium and, as already mentioned, may not correspond to the actual mechanism of the reaction.

Let us now consider the rate at which a reaction occurs. Individual molecules can react with one another when they collide, but not all collisions bring about reactions: in reality, usually only a very small fraction of all collisions result in reactions between molecules. The explanation of this is as follows.

In a reaction, the atoms of the colliding molecules are rearranged in a certain manner. For simplicity, let us assume that the reaction consists of a transfer of one atom from one molecule (A) to another molecule (B). The potential energy of this atom depends on its position with respect to the two molecules. The form of this energy as a function of a coordinate x along the "path of transition" of the atom is represented diagrammatically by a curve of the kind shown in Fig. 112. This diagram is, of course, highly schematic, since in reality the potential energy depends on several parameters (coordinates) and not on only one. What is important is not the precise variation of the potential energy but simply the fact that it has two minima corresponding to positions of the atom in each of the two molecules. These two positions are separated by a potential barrier.

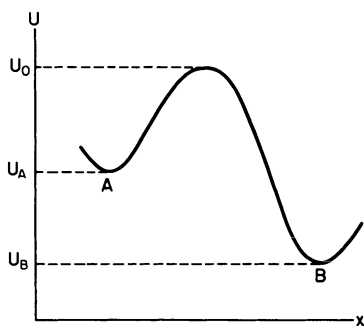


FIG. 112.

A chemical reaction can occur only if the atom which is to be transferred between two colliding molecules has sufficient energy to traverse the barrier. In the majority of molecules, however, this atom has an energy equal to or close to the corresponding minimum. The molecule can react, therefore, with the transition $A \rightarrow B$, only if it has excess energy equal to $U_0 - U_A$ (see Fig. 112). The ratio of the number of such molecules in the gas to the num-

ber which do not possess such energy is equal to the ratio of the Boltzmann factors (see §55):

$$e^{-U_0/kT} : e^{-U_A/kT} = e^{-(U_0-U_A)/kT}.$$

The energy $U_0 - U_A$ is called the *activation energy* of the reaction concerned, and usually referred to one mole of the substance by multiplying $U_0 - U_A$ by Avogadro's number: $N_0(U_0 - U_A) = E$.

Thus the number of molecules capable of reacting, and therefore the reaction rate, are proportional to the *activation factor*

$$e^{-E/RT}.$$

This is the principal factor in the temperature dependence of the reaction rate, and we see that the reaction rate increases very rapidly with temperature.

If the reaction rate is denoted by v , it follows from the above that

$$\log_e v = \text{constant} - E/RT,$$

i.e. the logarithm of the reaction rate is a linear function of $1/T$. The slope of the straight line representing this function gives the activation energy E .

The activation energy may have very different values for different molecular processes. For the majority of observable reactions it lies in the range from 10 to 150 kJ.

When the temperature changes from a value T to a slightly different value $T + \Delta T$, the change in the reaction rate is given by the formula

$$\begin{aligned} \log_e v_2 - \log_e v_1 &= \log_e(v_2/v_1) \\ &= -\frac{E}{R(T + \Delta T)} + \frac{E}{RT} \approx \frac{E}{RT^2} \Delta T. \end{aligned}$$

For example, when $E = 80$ kJ, $T = 300^\circ\text{K}$, $\Delta T = 10^\circ$, we find $v_2/v_1 \approx 3$. This is a typical increase in the reaction rate. For many kinds of reactions in gases and solutions, it is found that an increase of 10° in the temperature (in the range where the

reaction occurs at an appreciable rate) increases the reaction rate by a factor between two and four.

The extent to which the reaction rate depends on the temperature can be seen, for example, from the reaction $2\text{HI} \rightarrow \text{H}_2 + \text{I}_2$, whose activation energy is 185 kJ. At 200°C the reaction still hardly occurs at all: the dissociation of an appreciable quantity of HI would take hundreds of years. At 500°C, the reaction is complete within seconds; yet even at this temperature only about one in 10^{12} collisions between HI molecules results in dissociation.

The necessity for a sufficiently high energy of the molecules is the principal reason for the low efficiency of collisions in producing reactions. It is indispensable that the necessary excess energy should be concentrated on certain atoms or groups of atoms in the molecule; this fact also has a decisive influence in establishing the reaction rate. For reactions which involve complex molecules, there is also a geometrical factor: it is necessary that the reactive parts of colliding molecules should be in contact.

Let us again consider Fig. 112. The difference $U_A - U_B$ corresponds to the difference of the internal energies of molecules A and B , i.e. the heat of reaction evolved in the exothermic reaction $A \rightarrow B$ or absorbed in the endothermic reverse reaction $B \rightarrow A$. This difference is not directly related to the height of the potential barrier, i.e. there is no direct relation between the heat of reaction and the activation energy of the reaction. But there is a relation between the heat of reaction and the difference of the activation energies of the forward and reverse reactions. The diagram shows that the activation energies of the reactions $A \rightarrow B$ and $B \rightarrow A$ are $U_0 - U_A$ and $U_0 - U_B$; their difference is equal to the heat of reaction:

$$(U_0 - U_B) - (U_0 - U_A) = U_A - U_B.$$

As already mentioned in §88, reactions do not usually proceed exactly in accordance with the overall chemical equation: in reality, most chemical reactions have a more or less complex mechanism consisting of a number of simple elementary processes, the intermediate stages of the reaction, which are often difficult to determine. The reaction as it were selects the quickest path. The intermediate stages of the reaction must, of course,

have the lowest possible activation energies; this is the fundamental physical factor which determines the path of the reaction. The different stages may occur at very different rates. The rate of the overall process will evidently be determined mainly by the slowest of these intermediate stages, just as the speed of an assembly line can never be faster than the speed of the slowest operation.

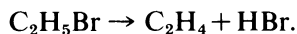
Decreasing the activation energy of the intermediate stages of a reaction is the basis of most processes of *catalysis*, i.e. the acceleration of reactions by adding to the reaction mixture another substance called a *catalyst*. This acceleration may be very great; reactions which otherwise practically do not occur at all often take place rapidly when a catalyst is present. The function of the catalyst is to participate in intermediate reactions in some way, while being restored to its original form as a result of the whole process.

It should be emphasised that a catalyst can not displace the position of chemical equilibrium, which does not depend on how the reaction occurs. The only effect of the catalyst is on the rate at which equilibrium is established.

§92. Molecularity of reactions

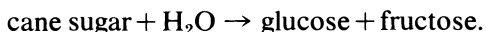
All chemical reactions in gases or in weak solutions can be assigned to a number of types, depending on the number of molecules which must collide in order to bring about the reaction. Here, it must be emphasised, the true molecular processes which actually occur are meant. In the examples given below the reactions in fact occur in the way shown by the chemical equation. In most cases, however, this classification of reactions relates to the individual elementary stages of a complex reaction mechanism.

Monomolecular reactions are those in which the molecules of the original substance decompose into two or more parts, for example the decomposition of ethyl bromide:



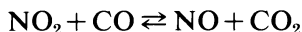
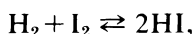
There is no need for molecules to collide in order to bring about such reactions. Thus, as the substance decomposes, the reaction rate decreases linearly with its concentration.

In this sense reactions in weak solutions behave similarly when solvent molecules participate in addition to one molecule of solute, for example in the hydrolysis of cane sugar already mentioned above:



This reaction in fact involves two molecules, but since there are plenty of water molecules round every sugar molecule throughout the reaction, the change in the reaction rate is due solely to the change in the concentration of dissolved sugar.

Reactions in which two molecules yield two or more molecules are said to be *bimolecular*; for example, the reactions



are bimolecular in both directions. A collision between two molecules is necessary for such reactions to occur, and the rate is therefore proportional to the product of concentrations of the reacting substances (or to the square of the concentration, if the reaction involves two identical molecules). This type comprises the great majority of the elementary processes which make up the mechanism of complex reactions.

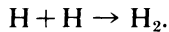
Finally, *trimolecular* reactions are those in which three molecules take part to give two or more other molecules. These are comparatively few, because they can occur only if three molecules collide simultaneously, and such ternary collisions are, of course, much rarer than collisions between pairs of molecules.

It is easy to determine the ratio of the numbers of ternary and binary collisions of molecules in a gas. We can say that ternary collisions of a given molecule are those which it undergoes while at the same time in the vicinity of a third molecule. Let V denote the total volume occupied by the gas, and b the total volume of the gas molecules. It is evident that the volume within which a molecule must be situated in order to be considered as in the vicinity of some other molecule is of the order of b , and the probability that the molecule is in the vicinity of another molecule is therefore b/V . The ratio of the numbers of ternary and binary collisions is

consequently also of the order of b/V . This is a small quantity; for example, for air under standard conditions it is about 10^{-3} .

The number of quaternary collisions is less than that of ternary collisions in the same ratio. Because of the extreme rarity of such collisions, chemical reactions of higher orders (quadrimeric, etc.) do not occur in Nature.

Some reactions which would appear to be bimolecular are in fact trimolecular. These are reactions in which two particles combine into one, for example



If an H_2 molecule were formed by a collision of two H atoms, it would immediately dissociate again; the two colliding atoms can always move apart again. A stable H_2 molecule must have a negative internal energy. Thus two hydrogen atoms can form a stable molecule only when a further particle is present to receive the excess energy liberated in the formation of the molecule. This means that the reaction in question actually occurs only in a collision between three particles.

It is interesting to note that even reactions which are clearly monomolecular sometimes behave as if they were bimolecular. In order to decompose, a molecule must have sufficient energy for its parts to overcome the potential barrier as they separate. An "activated" molecule of this kind has a definite "lifetime"; in a complex molecule, for example, the excess energy must be concentrated at the point where it is required for the decomposition. Activated molecules are formed as a result of collisions between molecules in their thermal motion. In a sufficiently rarefied gas, where collisions are comparatively infrequent, activated molecules decompose more rapidly than fresh ones are formed. Under these conditions, the reaction rate is mainly determined by the rate of the activation process, which requires collisions between molecules and is therefore a bimolecular process.

§93. Chain reactions

A characteristic feature of the mechanism of the majority of reactions is that fragments of molecules appear as intermediate products. These are individual atoms or groups of atoms, known as *free radicals*, which do not exist in a stable state.

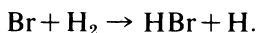
For example, in the decomposition of heated nitrous oxide gas (for which the formal equation is $2\text{N}_2\text{O} = 2\text{N}_2 + \text{O}_2$), the N_2O molecules decompose thus: $\text{N}_2\text{O} \rightarrow \text{N}_2 + \text{O}$, forming free atoms of oxygen, which then react with further molecules of nitrous oxide: $\text{O} + \text{N}_2\text{O} \rightarrow \text{N}_2 + \text{O}_2$.

In this instance the intermediate particles (O atoms) disappear after the two component processes have occurred. There are many reactions, however, in which active intermediate products react continuously, and thus act as a kind of catalyst.

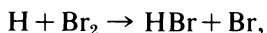
This very important type of reaction mechanism may be illustrated by the formation of hydrogen bromide in a mixture of hydrogen and bromine vapour when the mixture is exposed to the action of light. This reaction in fact does not occur by collision of H_2 and Br_2 molecules, as would correspond to the chemical equation $\text{H}_2 + \text{Br}_2 = 2\text{HBr}$. Its true mechanism is as follows. Under the action of light, Br_2 molecules dissociate into two atoms:



This is called *chain initiation*, and the bromine atoms formed act as *active centres*. These atoms, on colliding with H_2 molecules, react thus:



The resulting H atoms in turn react with Br_2 molecules:



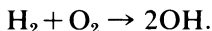
again forming bromine atoms, which react with H_2 molecules, and so on. A continuous chain of successive reactions results, in which the Br atoms act as a kind of catalyst, being restored unchanged after the formation of two HBr molecules. This is called a *chain reaction*. The principles of the theory of chain reactions were worked out by N. N. Semenov and C. N. Hinshelwood.

We see that, if active centres are formed in some way, the chain reaction will then proceed spontaneously, and could go to completion, one might think, without further external interaction. In reality, however, *chain termination* must also be taken into account. One active centre—a bromine atom in the above

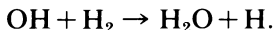
example—can cause the reaction of hundreds of thousands of hydrogen and bromine molecules, but it must eventually be lost, thus stopping the further progress of the chain reaction. This can occur, for example, by the recombination of two Br atoms to form a Br₂ molecule. It has been mentioned in §92, however, that such a combination of two atoms to form a stable molecule can occur only by a ternary collision. This mechanism of chain termination therefore becomes important only at high pressures, when ternary collisions in the gas are fairly frequent.

Another mechanism of chain termination consists in the loss of active centres when they strike the walls of the reaction vessel. This is of importance at low pressures, when the active centres can move quite easily through the gas.

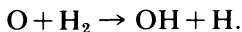
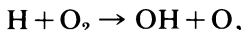
On the other hand, there exist reactions in which *chain branching* occurs. For example, the combustion of hydrogen in a detonating mixture of hydrogen and oxygen occurs (at high temperatures) essentially as follows. By an external interaction (e.g. passage of a spark) a chain is initiated:



The resulting active centres (OH radicals) react with H₂ molecules to give water:



The H atoms thus formed then react as follows:



These reactions not only yield water but also increase the number of active centres H, O and OH (unlike the reaction of HBr formation, where the number of free H and Br atoms did not increase).

If the increase in the number of active centres by chain branching outweighs the termination of chains, then this number grows very rapidly (in geometric progression), and the reaction is thereby accelerated into an *explosion*.

This *chain mechanism* of explosion has the feature that it can in principle develop even at constant temperature. Another important explosion mechanism is the *thermal mechanism*, which results from the marked temperature dependence of the reaction rate. When heat is rapidly evolved in an exothermic reaction, the rate of removal of heat may be insufficient, and consequently the reaction mixture will be heated, thus leading to a progressive spontaneous acceleration of the reaction.

SURFACE PHENOMENA

§94. Surface tension

So far we have discussed thermal properties and phenomena which occur throughout volumes and affect the whole mass of a body. The existence of free surfaces of bodies brings about the existence of a separate class of *surface phenomena* or *capillarity effects*.

Strictly speaking, any body is in an external medium, such as the atmosphere, and not in a vacuum. Thus we should speak not merely of the surfaces of bodies but of interfaces between two media.

In surface phenomena, only those molecules which are actually at the surfaces of bodies are involved. If the bodies are not very small, the number of such molecules is very small in comparison with the total number of molecules in the bodies. For this reason, surface phenomena are usually of minor importance, but they become significant in small bodies.

The molecules which are in a thin layer adjoining the surface are in conditions different from those within the body. The latter are surrounded by similar molecules on all sides, whereas the molecules near the surface have similar molecules on one side only. This has the result that the energy of the molecules in the surface layer is different from that of the molecules within the body. The difference between the energy of all the molecules (in both media) which are near the surface and the energy which they would have within the bodies is called the *surface energy*.

It is evident that the surface energy is proportional to the area S of the interface:

$$U_{\text{surf}} = \alpha S.$$

The coefficient α depends on the nature and state of the media in contact; it is called the *surface tension*.

As we know from mechanics, forces always act so as to bring a body to a state of minimum energy. In particular, the surface energy tends to take its least possible value. Hence it follows that the coefficient α is always positive, since otherwise the media in contact could not exist separately; their interface area would tend to increase without limit, i.e. the two media would tend to mix.

Conversely, since the surface tension is positive, the interface between two media must always tend to contract. This is the reason why droplets of liquid (and gas bubbles) tend to be spherical: for a given volume, the sphere is the figure with the least area. This tendency is opposed by the force of gravity, but for small droplets its effect is slight and their shape is almost spherical. In conditions of weightlessness this will be the shape of any free mass of liquid. Such conditions may be simulated in a well-known experiment with a spherical drop of oil floating within a mixture of alcohol and water having the same specific gravity.

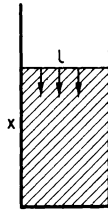


FIG. 113.

The surface-tension force is shown by the following simple example. Let us imagine a film of liquid supported on a wire frame of which one side (of length l) is movable (Fig. 113). Because the surface tends to contract, the wire is subject to a force which can be directly measured on the movable part of the frame. By the general laws of mechanics, this force is the derivative of the energy (in this case, of the surface energy) with respect to the coordinate x in the direction in which the force acts:

$$F = - dU_{\text{surf}}/dx = - \alpha dS/dx.$$

But the area of the film is $S = lx$, and therefore

$$F = -\alpha l.$$

This is the force on a segment l of the frame due to the surface-tension force on one side of the film; since the film has two sides, the force on the segment l is twice this value. The minus sign shows that this force is directed into the surface of the film.

Thus the line bounding the surface of the body (or any part of this surface) is subject to forces perpendicular to the line and tangential to the surface, directed into the surface. The force per unit length is equal to the surface tension α .

The dimensions of α follow from its definition, and may be put in various forms: energy per unit area, or force per unit length,

$$[\alpha] = \text{erg/cm}^2 = \text{dyn/cm}.$$

It is clear from the above that, in stating the value of the surface tension, it is necessary to state which two media are in contact. The term "surface tension" is often applied to a liquid (without specifying any other medium) to denote the surface tension between the liquid and its vapour. This quantity always decreases with increasing temperature and becomes zero at the critical point, where the difference between liquid and vapour ceases to exist.

The following list gives the surface tension (in erg/cm^2) between various liquids and air:

Water (20°C)	73	Mercury (20°C)	480
Ethyl ether (20°C)	17	Gold (1130°C)	1100
Benzene (20°C)	29		

Liquid helium has a very low surface tension at an interface with its vapour, only 0.35 erg/cm^2 (near absolute zero).

A surface tension also exists, of course, at the surfaces of solids, but here its effect is very slight under ordinary conditions: the comparatively weak surface forces cannot change the shape of a solid body. A direct measurement of the surface tension of solids is therefore very difficult, and there are no reliable data as to its values.

The surface tension of an anisotropic body (a crystal) must be different on different faces, since in general the atoms are differently arranged on different faces. For this reason, if a crystal could freely change shape under the action of external forces, it would not become spherical as an isotropic body (a liquid) would, where the surface tension is everywhere the same. It can be shown that the equilibrium shape of a crystal under these conditions is a very curious one, consisting of a relatively small number of plane faces, which, however, do not meet at angles but are joined by rounded regions.

This phenomenon may be observed, for example, on prolonged heating (at about 750°C) of spheres of rock salt cut from single crystals. The high temperature assists the atoms in “creeping” from one point on the surface to another, and in consequence the sphere is converted into a figure of the kind described.

§95. Adsorption

Many surface phenomena come under the heading of *adsorption*, which consists in the adhesion of substances on the surfaces of solids and liquids (the latter being then called *adsorbents*). Adsorption can take place from gases or liquids, and a solute may be adsorbed from solution. For example, many gases are adsorbed on the surface of carbon, silica gel, or the majority of metals; carbon adsorbs various organic compounds from solution. The degree of adsorption is described by the *surface concentration*, which is the quantity of the substance per unit area of the surface of the adsorbent.

Adsorption phenomena are widely found in Nature, and play an important part in technology. In order to adsorb a large quantity of a substance, we must evidently use substances which have the maximum area for a given mass, such as porous or finely powdered materials. To describe this property of adsorbents, we use the *specific area*, which is the area per unit mass of the substance. In good adsorbents, such as specially prepared porous carbons, it reaches hundreds of square metres per gram. Such large values of the specific area are not surprising if we consider how rapidly the surface area increases when a body is permeated by pores or is finely crushed. For example, 1 cm^3 of material in spheres of radius $r\text{ cm}$ will have a total area of $3/r\text{ cm}^2$, and when $r \sim 10^{-6}$ this amounts to hundreds of square metres.

The concentration of adsorbed gas depends (at a given temperature) on the gas pressure over the surface of the adsorbent. This dependence is shown by a curve, called an *adsorption isotherm*, of the form shown in Fig. 114. The surface concentration at first increases rapidly with pressure. As the pressure continues to rise, the concentration increases more slowly, and finally reaches a limit or saturation value. Experiment shows that the saturation of adsorption corresponds to a more or less dense occupation of the adsorbent surface by a single layer of adsorbed molecules (called a *monomolecular layer*).

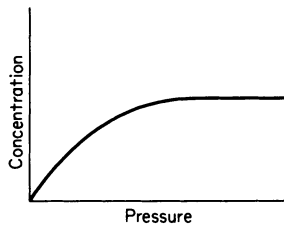


FIG. 114.

A very important property of adsorption is the change which it causes in the surface tension at the interface between media. Usually the surface of a liquid is concerned. Adsorption always reduces the surface tension, since otherwise adsorption would not occur. Here again there is a tendency to reduce the surface energy: this reduction can be achieved not only by decreasing the surface area but also by changing the physical properties of the surface. Because of their effect on the surface tension, substances which can be absorbed (on the surface of a given liquid) are said to be *surface-active*. On water, for example, various soaps are surface-active.

The total quantity of a substance which can be absorbed on the surface of a liquid is very small. Thus even small quantities of surface-active substances accumulating on the surface of a liquid may considerably affect its surface tension. The surface tension of a liquid is very sensitive to impurities: for example, even very small quantities of soap can reduce the surface tension of water by a factor of more than three.

Adsorbed monomolecular films on the surfaces of liquids are a very curious physical phenomenon, forming as it were a two-

dimensional state of matter, in which the molecules are distributed over a surface in two dimensions and not over a volume in three dimensions. In this state there can exist various phases, "gas", "liquid" and "solid", exactly analogous to three-dimensional phases.

In a "gaseous" film the adsorbed molecules have a comparatively rarefied distribution on the surface of the liquid and can move freely on it. In "liquid" and "solid" films the molecules are close together, either retaining some freedom of relative motion (so that a liquid film can "flow"), or so firmly held together that the film behaves as a solid. Liquid and solid films may be anisotropic, forming two-dimensional analogues of liquid and solid crystals; in the liquid film we have a regular orientation of molecules on the surface of the adsorbent, and in the solid film a type of two-dimensional crystal lattice with a regular configuration of molecules. It is noteworthy that such anisotropic films may occur at an interface between two isotropic media, a liquid and a gas.

These effects are very well illustrated by the monomolecular films formed on a water surface by insoluble complex organic acids, alcohols etc., whose molecules form long hydrocarbon chains with $-\text{COOH}$, $-\text{OH}$, etc., groups at one end. These groups are strongly attracted by the water molecules and, as it were, dissolve in the surface layer of the water, but cannot carry the whole molecule into the liquid; part of the molecule remains above the surface. Thus a liquid or solid film forms a kind of forest of closely packed molecules with their ends immersed in water.

The surface tension α when the water surface is covered with a film is less than its value α_0 for the clean surface. The difference $\alpha - \alpha_0$ can be measured directly from the force acting on a barrier floating freely on the surface of the water and separating the film from the clean surface. The film exerts a force α (into the film) per unit length of this barrier, and there is an opposite force α_0 exerted by the clean surface. Since $\alpha_0 > \alpha$, the result is that the film repels the barrier with a force

$$\Delta\alpha = \alpha_0 - \alpha$$

per unit length. This force may be regarded as the pressure of the film. At a given temperature, it is a definite function of the area S of the film (formed by a given quantity of the adsorbed

substance), just as the pressure of an ordinary body is a function of its volume.

For a gaseous, rarefied film (with n molecules in the area S), this relation is given by

$$\Delta\alpha = nkT/S,$$

which is similar to the equation of state of an ideal gas ($p = NkT/V$). When the film is compressed (i.e. when its area S decreases), there occurs at a certain value of $\Delta\alpha$ a phase transition to a continuous liquid or solid film. On the curve of $\Delta\alpha$ as a function of S this transition corresponds to a horizontal section, which is entirely similar to that for the ordinary transition between vapour and liquid on the isotherms which show the relation between the pressure p and the volume V (§70).

§96. Angle of contact

At the edge of a liquid surface in a vessel we have three media in contact: the solid wall (medium 1 in Fig. 115), the liquid (2) and the gas (3). Let us consider the capillary effects which occur at such a boundary.

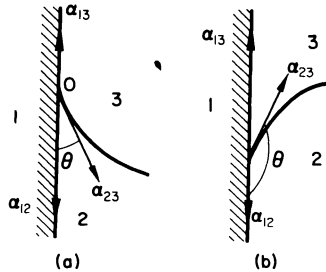


FIG. 115.

Three forces of surface tension act on the line along which all three media are in contact; this line intersects the plane of the diagram at O . Each force is directed tangentially inwards along the interface between the two media, as shown by the arrows in the diagram. The magnitudes of the forces per unit length of the line of contact are equal to the respective surface tensions α_{12} , α_{13} , α_{23} . The angle between the surface of the liquid and the solid wall will be denoted by θ and is called the *angle of contact*.

The surface of the liquid takes a form such that the resultant of the three forces α_{12} , α_{13} , α_{23} has no component along the wall of the vessel (the component perpendicular to the wall is balanced by the reaction of the wall). Thus the condition of equilibrium of the liquid at the wall is

$$\alpha_{13} = \alpha_{12} + \alpha_{23} \cos \theta,$$

whence

$$\cos \theta = (\alpha_{13} - \alpha_{12}) / \alpha_{23}.$$

We see that the angle of contact depends only on the nature of the three media in contact (through the surface tensions at their interfaces); it does not depend on the shape of the vessel or on the force of gravity acting on the bodies. It must be remembered, however, that the surface tensions, and therefore the angle of contact, are very sensitive to the state of cleanliness of the interfaces.

If $\alpha_{13} > \alpha_{12}$, i.e. if the surface tension at the interface between the solid wall and the gas is greater than that at the interface between the wall and the liquid, then $\cos \theta > 0$ and the angle θ is acute. In other words, the edge of the liquid is raised, and its surface or *meniscus* is concave (Fig. 115a). The liquid is then said to *wet* the solid surface. If a drop of such a liquid is placed on the surface of the solid, it “flows” to some extent over the surface (Fig. 116a).

If, on the other hand, $\alpha_{13} < \alpha_{12}$, then $\cos \theta < 0$ and θ is obtuse; the edge of the liquid is depressed and its meniscus is convex (Fig. 115b). In this case we say that the liquid *does not wet* the solid. For example, the angle of contact of mercury on glass is about 150° , and that of water on paraffin wax is about 105° . Drops of such liquids, when placed on the solid surface, appear to contract so as to reduce the area of contact with the surface (Fig. 116b).



FIG. 116.

Since the cosine of an angle cannot exceed unity in absolute magnitude, it is seen from the formula derived above for $\cos \theta$ that in any actual case of stable equilibrium between the liquid and the wall the condition

$$|\alpha_{13} - \alpha_{12}| \leq \alpha_{23}$$

must hold. If α_{12} , α_{13} , α_{23} are taken to be the surface tensions for each pair of media alone, in the absence of the third medium, this inequality may certainly prove to be violated. In reality, we must remember, the third substance may be adsorbed on the interface between the other two and thus lower the surface tension, so that the resulting values of α are such as to satisfy the foregoing condition.

The concepts of wetting and non-wetting in the sense explained above must be distinguished from that of *complete wetting*, which refers to the condensation of a vapour on the surface of a solid. As we know, the condensation of a vapour to a liquid is brought about by the action of the van der Waals forces of attraction between molecules. These forces, however, can be exerted on a molecule in the vapour not only by similar molecules but also by the molecules of a solid. Let us suppose that the attraction forces from the solid are stronger than those in the liquid itself. In such a case, the presence of the solid surface will clearly bring about a partial condensation of the vapour even in conditions where the vapour is unsaturated and would therefore otherwise be stable. A thin film of liquid forms on the surface of the solid. The thickness of such a film cannot be great, of course; its order of magnitude is determined by the range of action of the van der Waals forces and may be from 10^{-7} to 10^{-5} cm. As the vapour approaches saturation the film becomes thicker. This effect is called complete wetting of the solid by the liquid. For example, carbon tetrachloride (CCl_4) completely wets many surfaces, including that of glass.

[The difference between this phenomenon and adsorption should be emphasised: here we are discussing a very thin but still "macroscopic" layer of liquid, whereas an adsorbed film consists of individual molecules distributed over the surface.]

The edge of a liquid which completely wets the walls of a vessel passes continuously into the film on the wall. Thus in this case

there is no finite angle of contact. We may say that complete wetting corresponds to zero angle of contact. A drop of such a liquid placed on the surface will spread completely over it.

More complex types of wetting are in principle possible, depending on the nature of the van der Waals forces exerted by the solid. For instance, a case is possible where the vapour condenses into a liquid on the solid surface but the thickness of the resulting film can not exceed a certain limiting value. If the surface is already covered with such a film, a further drop of liquid placed on it will not spread completely, but will remain isolated, though highly flattened, with a very small but finite angle of contact. This seems to occur for water on clean glass; the maximum film thickness is about 10^{-6} cm, and the angle of contact is probably less than one degree.

§97. Capillary forces

It has several times been mentioned that, in a state of equilibrium, the pressures of bodies in contact must be equal. In reality this statement is true only in so far as capillary effects are neglected. When the surface tension is taken into account, the pressures in adjoining media are in general different.

Let us consider, for example, a drop of liquid in air. The tendency of the drop surface to decrease causes a contraction of the drop and therefore an increase in its internal pressure. The pressure of the liquid in the drop therefore exceeds the pressure of the surrounding air. The difference between them is called the *surface pressure* and will be denoted by p_{surf} .

To calculate this quantity, we note that the work done by the surface forces in reducing the surface area of the drop by dS is equal to the corresponding decrease αdS in the surface energy. This work can also be written as $p_{\text{surf}}dV$, where dV is the change in the volume of the drop; thus

$$\alpha dS = p_{\text{surf}}dV.$$

For a spherical drop of radius r , $S = 4\pi r^2$ and $V = 4\pi r^3/3$; substitution in the above equation then gives the following expression for the surface pressure:

$$p_{\text{surf}} = 2\alpha/r.$$

This formula applies also, of course, to a bubble of gas in a liquid. The higher pressure always occurs in the medium towards which the interface is concave. When $r \rightarrow \infty$, the surface pressure tends to zero. This is in accordance with the fact that for a plane interface the pressures in the adjoining media must be the same; it is evident that the tendency of the surface to contract will not lead to any force into either medium in this case.

We may also derive a formula for the surface pressure in a cylindrical mass of liquid. In this case $S = 2\pi rh$, $V = \pi r^2 h$ (where r is the radius and h the height of the cylinder), and substitution in the equation $p_{\text{surf}} dV = \alpha dS$ gives

$$p_{\text{surf}} = \alpha/r.$$

These simple formulae enable us to solve a number of problems relating to capillarity effects.

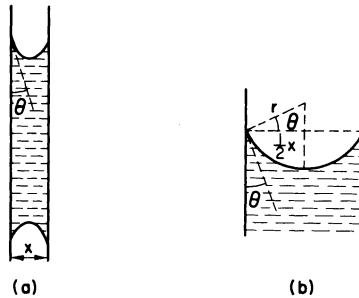


FIG. 117.

Let us imagine two parallel flat plates (shown in cross-section in Fig. 117) between which is a thin layer of liquid. The lateral surface of the liquid is in contact with air. If the angle of contact is acute, the meniscus of the liquid is concave and the pressure within the liquid is less than the air pressure; the atmospheric pressure acting on the plates will therefore tend to bring them together and they appear to attract each other; if the angle of contact is obtuse, the meniscus is convex and the layer of liquid appears to push the plates apart. When the space between the plates is sufficiently narrow, any small section of the meniscus may be regarded as part of a cylindrical surface of some radius r .

A simple construction (Fig. 117b) shows that $x = 2r \cos \theta$, where x is the distance between the plates. The pressure in the liquid is less by $p_{\text{surf}} = \alpha/r = (2\alpha/x) \cos \theta$, where α is the surface tension between the liquid and the air. The force of attraction F between the plates is found by multiplying this quantity by the area of contact S between the liquid and each plate:

$$F = (2\alpha S/x) \cos \theta.$$

We see that this force is inversely proportional to the distance between the plates. When the distance is small, the force may be very large; for example, plates separated by a film of water one micron thick are attracted together by a pressure of about 1.5 atm.

Let us next consider the well-known *capillary rise* (or fall) of a liquid in a narrow tube immersed in the liquid. When the meniscus is concave (acute angle of contact) the pressure of the liquid in the tube is less than that of the adjacent air by an amount p_{surf} . The atmospheric pressure on the surface of the liquid in the vessel therefore causes the level of the liquid in the tube to rise until the weight of the column of liquid balances the extra pressure: $p_{\text{surf}} = \rho gh$, where ρ is the density of the liquid. The surface of the meniscus in a narrow tube may be regarded as part of a sphere whose radius r is related to the radius a of the tube by $a = r \cos \theta$. Then $p_{\text{surf}} = 2\alpha/r = (2\alpha/a) \cos \theta$ and the height to which the liquid rises is

$$\begin{aligned} h &= 2\alpha/g\rho r \\ &= (2\alpha/g\rho a) \cos \theta. \end{aligned}$$

For a convex meniscus, this formula gives the depth to which the liquid sinks.

The surface tension and the density of the liquid appear in the above formula in the combination $\alpha/\rho g$. The quantity $\sqrt{2\alpha/\rho g}$ has the dimensions of length and is called the *capillary constant*. It plays an important part in all phenomena which occur under the combined action of surface-tension forces and gravity. The capillary constant of water at 20°C is 0.39 cm.

Various effects of capillary forces are the basis of methods for measuring surface tension. For example, the size of drops of

liquid flowing slowly from a narrow tube is determined by the equilibrium between the weight of the drop and the surface tension around its "neck"; thus a measurement of the weight of the drop (by counting the number of drops formed by a given quantity of liquid) enables us to determine α . Another method is based on measurement of the surface pressure within a gas bubble of given radius; this is effected by measuring the additional pressure necessary to expel a bubble of air from the end of a tube immersed in the liquid.

§98. Vapour pressure over a curved surface

The influence of capillary forces causes some changes also in the properties of equilibrium between a liquid and its saturated vapour. It has been stated above that the saturated vapour pressure is a definite function of temperature. In reality, this pressure depends also on the shape of the liquid surface above which the vapour is situated. The dependence is admittedly very slight, and can be of importance only for small bodies (e.g. droplets of liquid).

The nature and amount of this dependence are easily determined by again considering the capillary rise (or fall) of a liquid and supposing that the space above the liquid in the vessel and in the tube is filled with saturated vapour. Since the gas pressure decreases with increasing height, it will evidently be smaller above a liquid which has risen (and larger above a liquid which has fallen) than above the flat surface of the liquid in the vessel. Comparing this with the shape of the meniscus in the tube in the two cases, we conclude that the saturated vapour pressure above a concave liquid surface is less (and above a convex surface greater) than above a flat surface. The similarity of this argument to the derivation of Raoult's law in §81 should be noted.

If h is the height of the capillary rise, the decrease in the saturated vapour pressure is $\Delta p = \rho_{\text{vap}}gh$. We have seen in §97 that $h = 2\alpha/\rho_l r g$, where ρ_l is the density of the liquid and r the radius of the sphere of which the meniscus forms part. Thus we have

$$\Delta p = \frac{2\alpha}{r} \frac{\rho_{\text{vap}}}{\rho_l}.$$

The decrease in the saturated vapour pressure above a concave surface causes what is called *capillary condensation*, the deposition of a liquid in a porous body from a vapour which under ordinary conditions would not be saturated. If the liquid wets the body concerned, concave menisci of liquid are formed in the pores (which act as very fine capillaries), and the vapour may then be supersaturated even at a comparatively low pressure.

When the liquid surface is convex, the same formula for Δp gives the amount by which the vapour pressure exceeds its value over a flat surface. We see that the saturated vapour pressure above a drop of liquid increases with decreasing radius of the drop.

Let us imagine a vapour containing a large number of liquid droplets of various sizes. It may happen that the vapour is supersaturated with respect to the larger droplets but unsaturated with respect to the smaller ones. Then the liquid which evaporates from the smaller drops will condense on the larger ones, which as it were “consume” the small drops.

§99. The nature of superheating and supercooling

The most important consequence of the dependence of the saturated vapour pressure on the size of a drop is that it gives an explanation of supersaturation of a vapour—the continued existence of the gaseous state under conditions such that the substance should become liquid.

A supersaturated vapour over the surface of the liquid will, of course, condense immediately, but if the vapour is not in contact with the liquid the condensation is impeded by the fact that it must begin with formation of small droplets in the vapour. A vapour supersaturated with respect to a flat liquid surface may still be unsaturated with respect to such droplets, which are then unstable and evaporate again as soon as they are formed. Only if a liquid drop happens to be formed in the vapour which is so large that the vapour is also supersaturated with respect to the drop will such a drop continue to exist and the vapour continue to condense on it; the drop will act as a *nucleus* of the new phase. Spontaneous formation of such nuclei in completely pure vapour can occur only by random thermal fluctuations, and this is in general a very unlikely occurrence. Its probability decreases for increasing values of the “critical”

radius of the drop, i.e. the minimum radius which gives stability. As the degree of supersaturation increases, the "critical" radius becomes smaller and the formation of nuclei becomes easier. When this quantity reaches values of the order of molecular dimensions, the creation of special nuclei is essentially unnecessary and further supersaturation of the vapour is impossible.

The condensation of a supersaturated vapour is assisted by the presence of a solid surface in contact with it which is wetted by the liquid in question. Small droplets which are deposited on such a surface spread somewhat and their surfaces become less curved. Thus such drops can easily become centres of further condensation. Condensation occurs with particular ease on a surface which is completely wetted by the liquid, since the drops disperse over the whole of such a surface.

Under ordinary conditions the vapour is not completely pure, and various small dust particles present in it act as centres of condensation, by forming solid surfaces which are wetted by the liquid. Thus, in order to achieve a considerable degree of supersaturation, a careful removal of all contamination from the vapour is necessary.

Charged particles (ions) strongly attract the vapour molecules, and consequently small droplets immediately form around them and act as centres for further condensation; thus charged particles create particularly favourable conditions for the condensation of the vapour. This phenomenon is, in particular, the basis of the *cloud chamber* used for the observation of the paths of fast ionising atomic or nuclear particles.

We have given a detailed discussion of the reasons for the occurrence of the metastable state of supersaturated (supercooled) vapour. These reasons are in fact general and account also for the "delay" in other phase transitions. The formation of a new phase within a previously existing one must begin with the formation of small inclusions or nuclei of the new phase. For example, the conversion of a liquid into a vapour must begin with the appearance in the liquid of small bubbles of vapour; the solidification of a liquid, with the appearance in it of crystal nuclei, and so on.

The additional surface energy at the boundary of such a nucleus makes its formation energetically unfavourable unless it is sufficiently large. Here we have a competition between two opposing

factors. The formation of a new interface between two phases involves the absorption of the surface energy, but when the substance enters a new phase there is a gain in volume energy. The latter quantity increases with increasing size of the nucleus more rapidly than the former quantity, and ultimately outweighs this. We may say that the formation of a nucleus of a new phase requires the traversing of a "potential barrier" due to the surface energy, and this is possible only for a sufficiently large nucleus.

There is one phase transition which in this respect appears to form an exception to the general rule, namely the melting of crystals. When crystals are heated in an ordinary manner, superheating is never observed. This, however, is simply because the surface of any crystal is completely wetted by the liquid formed when it melts. Thus the liquid droplets formed on the surface of the crystal spread over it, and surface tension does not act to prevent melting.

Superheating of crystals can occur if the crystal is artificially heated from the inside instead of from the outside. For instance, when a current is passed through a single-crystal rod of tin with intensive air cooling of the exterior, the temperature within the rod is higher than that on its surface, and the interior of the crystal can then be superheated by one or two degrees before ordinary melting begins at its surface.

§100. Colloidal solutions

Sometimes a substance which does not dissolve in a given liquid can be distributed in it in the form of very fine particles, although these still contain a very large number of molecules. In this case the finely divided or *dispersed* substance is called a *disperse phase*, and the medium in which it is distributed is called the *dispersion medium*. If the size of the particles is of the order of 10^{-4} to 10^{-2} cm, such a mixture is called a *suspension* or *emulsion* according as the particles are of a solid or a liquid; for example, milk is an emulsion of fat in water.

When the particles are even smaller (10^{-7} to 10^{-5} cm, or 10 to 10^3 Å), the mixture is called a *colloidal solution* or *sol*. The characteristic property of these solutions is the size of the particles of the disperse phase rather than the number of molecules in each particle. For example, in a colloidal solution of gold in water, each particle is of size 100–500 Å and contains millions

of gold atoms, but in solutions of such complex substances as proteins each colloid particle may contain only one molecule.

The dispersion medium may be either a liquid or a gas. For example, colloidal solutions in air (*aerosols*) may be smokes, mists or fogs. The most important colloidal solutions, however, are those in liquids, in particular those in water (*hydrosols*). For example, the majority of substances concerned in the constitution of plants and animals are present in them as liquid colloidal solutions.

Many kinds of substance are able to form sols: many organic compounds with large molecules (proteins, starch, gelatine, etc.), silicic acids, aluminium hydroxide, etc. Sols of some metals can also be obtained, for example of gold in water.

Because of the high degree of dispersion of the disperse phase, the total surface area of its particles is extremely large, and surface phenomena therefore have a very important effect on the properties of colloidal solutions.

Since the surface tension tends to reduce the area of the interface, the particles of the disperse phase have a tendency to combine and be precipitated from the solution as a dense mass. This tendency is counteracted by the forces of electrical repulsion: the particles of the disperse phase in a colloidal solution are always electrically charged, and all the charges are of the same sign (which may be either positive or negative). Only this fact prevents the particles from coalescing and being precipitated.

Colloidal particles are charged either because of the electrolytic dissociation of their molecules or by adsorption of ions from the surrounding fluid. When an electrolyte is added to a colloidal solution, the ions of the electrolyte may cancel the charge on the colloidal particles and render them electrically neutral. This brings about the precipitation or *coagulation* of the colloidal solution. Coagulation of colloids may also be effected by other means, for instance by heating.

Colloidal solutions may be divided into two groups as regards stability. Some colloidal solutions are a stable state of matter and can be precipitated only with difficulty. These are called *lyophilic* colloids, and include hydrosols of proteins, gelatine, silicic acids, and other substances. When a lyophilic colloidal solution is coagulated it often becomes a jelly-like mass called a *gel*. This contains not only the substance from the disperse

phase but also a considerable quantity of the solvent (water, etc.). A gel is a kind of irregular network of solute particles, enclosing solvent molecules. A typical feature of the conversion of a lyophilic sol into a gel is that it is reversible: under appropriate conditions a gel may absorb a sufficient quantity of solvent to become a sol again.

Colloidal solutions of the other group form a metastable state of matter and are very easily precipitated. These *lyophobic* colloids include, for example, colloidal solutions of metals in water. The coagulation of lyophobic colloids is accompanied by the formation of a dense precipitate, and is an irreversible process; the precipitate can not be so easily converted into a solution again.

CHAPTER XIII

MECHANICAL PROPERTIES OF SOLIDS

§101. Extension

The work done on a liquid or gas depends only on the change in its volume, and not on the change in shape of the vessel containing it. Liquids resist change in volume but not change in shape. This property is the reason for *Pascal's law* in liquids, which states that the pressure transmitted by a liquid is the same in all directions: if, for instance, a liquid is compressed by a piston, the same pressure will be exerted by the liquid on every wall of the vessel. The pressure force acting on the liquid and transmitted by the liquid is always at right angles to the walls: a force tangential to the surface, and not capable of being compensated because a liquid offers no resistance to a change of shape, can not exist in equilibrium conditions.

Solids, on the other hand, resist both change in volume and change in shape; they resist, therefore, any *deformation*. Work must be done even in order to change the shape alone of a solid, without altering its volume. We may say that the internal energy of a solid depends on its shape as well as on its volume. In consequence, Pascal's law does not apply to solids: the pressure transmitted by a solid is different in different directions. The pressures which occur in a solid when it is deformed are called *elastic stresses*. Unlike the pressure in a liquid, the elastic-stress force in a solid may be in any direction relative to the area on which it acts.

The simplest type of deformation of a solid is *extension*. This occurs in a thin rod (Fig. 118a) of which one end is fixed, when a force F tending to stretch the rod is applied to the other end. (If the force F is in the opposite direction we have a *compression*.) It may be noted that fixing in a wall is, by the law of action and

reaction, equivalent to applying to the fixed end a force equal and opposite to the force acting on the free end (Fig. 118b).

The elastic stresses in the rod are determined by the value F/S of the extending force per unit area of the cross-section S of the rod; let this value be p . The stresses are clearly constant along the length of the rod, and thus the same stretching stress p is exerted on each element of length of the rod by the adjoining parts of the rod (Fig. 118b). It is therefore clear that each unit

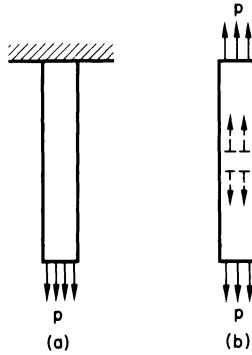


FIG. 118.

length of the rod undergoes the same extension, and the total increase δl in the length of the rod is proportional to this length. Thus the relative elongation

$$\lambda = \delta l / l_0$$

(where l_0 is the length of the rod before deformation) is independent of the length of the rod, and is clearly a measure of the degree of deformation undergone by each part of the rod.

Because of the high strength of solids, the deformations which they undergo when subjected to external forces are usually small. That is, the relative changes in size of solid bodies are small, and in the case of the extension described above the relative elongation is small. Such deformations may be assumed to be proportional to the stresses which cause them, and therefore to the magnitude of the external forces applied. This is called *Hooke's law*.

For extension, Hooke's law implies that the relative elongation

λ is proportional to the tensile stress p . This relation is usually written

$$\lambda = p/E,$$

where the coefficient E is a property of the material and is called *Young's modulus*. The relative elongation λ is evidently a dimensionless quantity, and the modulus E therefore has the dimensions of p , i.e. those of pressure.

As examples, the values of Young's modulus (in millions of bars) for a number of materials are as follows:

Iridium	5.2	Quartz	0.73
Steel	2.0-2.1	Lead	0.16
Copper	1.3	Ice (-2°C)	0.03

Young's modulus, however, does not completely describe the properties of a body with respect to deformation (its *elastic properties*). This is clear even for an extension. The reason is that longitudinal stretching of a rod involves a decrease in its transverse dimensions: the rod becomes thinner at the same time as its length increases. The value of Young's modulus enables us to calculate the relative elongation of the rod (for a given stress), but does not suffice to determine the transverse contraction.

The relative decrease in the transverse dimensions of the rod is also proportional to the tensile stress p , and therefore to the relative extension λ . The ratio of the relative transverse contraction of the rod to its relative elongation is a quantity characteristic of any given material and is called *Poisson's ratio*, denoted by σ . Thus the relative transverse contraction (e.g. the relative decrease in the diameter of a stretched wire) is

$$\sigma\lambda = \sigma p/E.$$

We shall see below that Poisson's ratio cannot exceed $\frac{1}{2}$. For most materials its value is in the range from 0.25 to 0.5. The value $\sigma = 0$ is reached in porous materials (such as cork) whose transverse dimensions are unaffected by stretching.

Thus the elastic properties of a solid are described by two quantities, E and σ . It should be emphasised, however, that we

have tacitly assumed the solid to be isotropic (the materials concerned are usually polycrystalline). The deformation of an anisotropic body (a single crystal) depends not only on the position of the external forces with respect to the body but also on the position of the crystallographic axes within the body. The elastic properties of crystals are of course described by a larger number of quantities than for isotropic bodies. The number increases with decreasing symmetry of the crystal, from 3 for cubic crystals to 21 for crystals of the triclinic system.

The work done on a body undergoing deformation is stored in the body in the form of *elastic energy*. Let us calculate this energy for a stretched rod. The work done by the tensile force F to increase the length of the rod by an infinitesimal amount $d(l_0\lambda) = l_0d\lambda$ is

$$dU = Fl_0d\lambda,$$

and this is also the increment of elastic energy. Substituting $F = Sp$, $p = E\lambda$, and noting that the product Sl_0 is the volume V of the rod, we obtain

$$SE\lambda.l_0d\lambda = VE\lambda d\lambda = VE d(\frac{1}{2}\lambda^2).$$

Hence it follows that, if the relative elongation of the rod changes from zero to some value λ , the work done is $\frac{1}{2}VE\lambda^2$. Thus the elastic energy per unit volume of the deformed rod is

$$U = \frac{1}{2}E\lambda^2,$$

which is proportional to the square of the deformation. This can also be put in the form

$$U = \frac{1}{2}\lambda p = p^2/2E.$$

An extension is a *uniform* deformation, i.e. one in which each volume element in the body is deformed in the same way. The *bending* of a thin rod is closely related to a simple extension (or compression), but is not a uniform deformation. Its nature is easily ascertained by imagining a rod bent into a circle. Before being bent the rod is straight, and so the length of each "fibre"

in it from one end to the other is the same. After the bending this is no longer true. The length of each fibre is $2\pi r$, where r is the radius of the circle which it forms, and this radius is less along the inner side of the rod than along the outer side. It is therefore clear that the inner part of the rod is compressed and the outer part is stretched. Since no lateral force is applied to the rod surface, the elastic stresses in the rod act only lengthwise, and this means that, in bending, each volume element is subjected to a simple extension or compression, though this is not the same for different elements: the parts nearer to the convex side of the bent rod are stretched, and those nearer the concave side are compressed.

§102. Uniform compression

The formulae for a simple extension are easily generalised to any uniform deformations.

Let a solid block in the form of a rectangular parallelepiped be stretched (or compressed) by forces acting on all sides and uniformly distributed over each face (Fig. 119). These forces

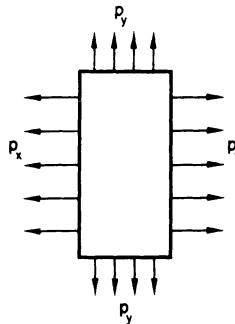


FIG. 119.

create elastic stresses in the body, which are in general different in three mutually perpendicular directions (along the three edges of the parallelepiped); let these stresses be p_x , p_y , p_z , with positive signs for tensile stresses and negative signs for compressive stresses. The relative changes in length in these directions (positive in extension and negative in compression) will be denoted by λ_x , λ_y , λ_z .

Let us consider this deformation as the result of three successive simple extensions along the three axes. For example, when stretched by the stress p_x the body is elongated in the x direction and shortened in the transverse y and z directions, with

$$\lambda_x = p_x/E, \quad \lambda_y = \lambda_z = -\sigma\lambda_x = -\sigma p_x/E.$$

Summation of the results of three such deformations gives

$$\lambda_x = \frac{p_x - \sigma(p_y + p_z)}{E}, \quad \lambda_y = \frac{p_y - \sigma(p_x + p_z)}{E}, \quad \lambda_z = \frac{p_z - \sigma(p_x + p_y)}{E}.$$

Next, let us find the change in the volume of the body as a result of the deformation. The volume of a parallelepiped with edges l_x, l_y, l_z is $V = l_x l_y l_z$. Taking logarithms, we find

$$\log_e V = \log_e l_x + \log_e l_y + \log_e l_z$$

and on differentiating

$$\frac{\delta V}{V} = \frac{\delta l_x}{l_x} + \frac{\delta l_y}{l_y} + \frac{\delta l_z}{l_z}.$$

The three terms in this sum are the relative elongations along the respective axes. Hence

$$\delta V/V = \lambda_x + \lambda_y + \lambda_z,$$

i.e. the relative change in volume is equal to the sum of the relative elongations in three mutually perpendicular directions.

Substitution of the expressions found above for $\lambda_x, \lambda_y, \lambda_z$ gives

$$\frac{\delta V}{V} = \frac{1 - 2\sigma}{E}(p_x + p_y + p_z).$$

Let us now consider some important particular cases of uniform deformation. If a body is subject to tensile (or compressive) stresses which are uniform in all directions, i.e. if the elastic stresses in it are the same in all directions ($p_x = p_y = p_z$), then the relative change in each dimension of the body is the same

($\lambda_x = \lambda_y = \lambda_z \equiv \lambda$). Such a deformation is called a *uniform extension* (or compression); in this case

$$\lambda = (1 - 2\sigma) p/E,$$

and the relative change in volume is

$$\delta V/V = 3\lambda = p/K,$$

where the coefficient

$$K = E/3(1 - 2\sigma)$$

is called the *modulus of uniform compression* or *bulk modulus*. Its reciprocal $1/K$ is clearly equal to the compressibility

$$\kappa = \frac{1}{V} \left| \frac{dV}{dp} \right|$$

discussed in §58. Thus the formula obtained relates the ordinary compressibility of a solid to the values of Young's modulus and Poisson's ratio.

The elastic energy stored in the body (per unit volume) in uniform compression is

$$U = \frac{1}{2}(\lambda_x p_x + \lambda_y p_y + \lambda_z p_z) = \frac{3}{2}\lambda p = \frac{1}{2}K\lambda^2 = \frac{1}{2}p^2/K.$$

The quantity K must always be positive, i.e. the volume of a body must be increased by extension and decreased by compression. It has been mentioned in §70 that bodies with the opposite dependence of the volume on the pressure would be absolutely unstable and therefore cannot exist in Nature. [This is also seen from the above formula for the elastic energy: if $K < 0$, this energy would be negative, and since a mechanical system tends towards the state of least potential energy, such a body would spontaneously undergo an unlimited deformation.]

Since K is positive, it follows that $1 - 2\sigma > 0$, or

$$\sigma < \frac{1}{2},$$

i.e. Poisson's ratio cannot exceed $\frac{1}{2}$.

Let us now consider the compression of a block held by lateral walls in such a way that its transverse dimensions may be regarded as constant (Fig. 120); this process is called *unilateral compression*.

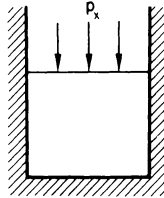


FIG. 120.

Let the direction of compression be along the x axis. The reaction of the walls which prevents a lateral expansion of the block gives rise to transverse stresses p_y and p_z in it. The magnitude of these is determined by the condition that the dimensions of the block in the y and z directions must remain unchanged ($\lambda_y = \lambda_z = 0$), and from symmetry we must have $p_y = p_z$. From the equation

$$\lambda_y = \frac{p_y - \sigma(p_x + p_z)}{E} = \frac{p_y(1 - \sigma) - \sigma p_x}{E} = 0$$

we find that the transverse stresses are related to the compressive pressure p_x by

$$p_y = p_z = \frac{\sigma}{1 - \sigma} p_x.$$

The longitudinal compression of the block is given by

$$\lambda_x = \frac{p_x - \sigma(p_y + p_z)}{E} = \frac{1 - \sigma - 2\sigma^2}{E(1 - \sigma)} p_x.$$

§103. Shear

Under uniform compression, the shape of a body remains the same, and only its volume changes. Deformations of the opposite kind are also of importance, where only the shape of the body changes and not its volume. These are described as *shear* deformations.

Since the volume is constant, we have

$$\delta V/V = \lambda_x + \lambda_y + \lambda_z = 0,$$

and hence

$$p_x + p_y + p_z = 0.$$

Substituting $p_y + p_z = -p_x$ in the formula

$$\lambda_x = \frac{p_x - \sigma(p_y + p_z)}{E},$$

we find that the relative elongation (or shortening) along any edge of a block is related to the stress in that direction by the formula

$$\lambda_x = \frac{1 + \sigma}{E} p_x.$$

This relation involves the quantity $E/(1 + \sigma)$; a quantity equal to one-half of this is called the *shear modulus* (or *modulus of rigidity*) and denoted by G :

$$G = E/2(1 + \sigma).$$

A shear deformation is, however, most simply brought about by applying to the block forces which are tangential, not perpendicular, to its surface. Let the lower face of the block be held fixed, and forces be applied in the plane of the upper face; stresses in this direction are often called *shearing stresses*. Under the action of these forces the parallelepiped becomes oblique, as shown in Fig. 121. The angle β (called the *angle of shear*) is small for small deformations (the only ones considered here). In a first approximation we can assume that the height of the parallelepiped is unchanged, and therefore that the volume is unchanged, giving a shear deformation. It can be shown that the angle β is related to the shearing force p (per unit area) by

$$\beta = p/G.$$

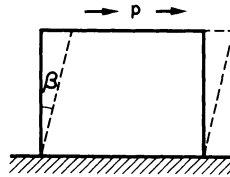


FIG. 121.

Like the modulus of uniform compression, the shear modulus must be positive, since the elastic energy that is stored in a body subjected to a shear deformation is positive only in that case. Hence it follows that we must have $1 + \sigma > 0$, i.e. $\sigma > -1$.

Using also the inequality $\sigma < \frac{1}{2}$ derived in §102, we can say that the values of Poisson's ratio for all bodies must lie in the range

$$-1 < \sigma < \frac{1}{2}.$$

These are the only conditions which follow from the general requirements of mechanical stability of solids. Thus in principle bodies could exist with negative values of σ . A rod of such a material should become wider in a simple extension, and not narrower as was assumed in §101. No bodies having such properties are known to exist in Nature, however, so that Poisson's factor in practice varies only between 0 and $\frac{1}{2}$. Values close to $\frac{1}{2}$ occur in substances such as rubber, which change their shape considerably more easily than their volume: their moduli of compression are large in comparison with their shear moduli.

The shearing of a rectangular block discussed above is a uniform deformation. The *torsion* of a rod is a pure shear but one which is not uniform. This occurs when one end of a rod is fixed and the other end is twisted. Different cross-sections of the rod are turned through different angles relative to the fixed base. Since neither the height nor the cross-sectional area of the rod is changed, its volume also remains constant.

It is easy to see how the shear deformation in torsion is distributed over the volume of the rod. Let us consider a rod of circular cross-section with radius R , and let its upper end turn through some angle ϕ relative to the lower end (Fig. 122). Any

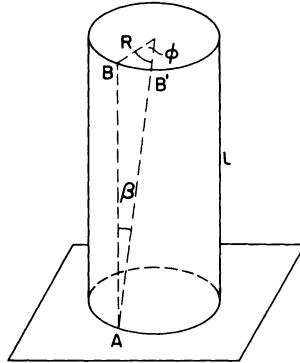


FIG. 122.

generator AB of the cylindrical surface of the rod then moves to the oblique position AB' . Since the distance $BB' = R\phi$, the small angle of shear β on the surface of the rod is

$$\beta \approx \tan \beta = R\phi/l,$$

where l is the length of the rod. Applying the same reasoning to a cylindrical surface of radius $r < R$, we find that it is likewise sheared, but through an angle

$$\beta_r = r\phi/l,$$

which is less than the angle of shear β at the surface of the rod. Thus in torsion the different elements of the rod undergo different degrees of shearing, which become smaller as the axis of the rod is approached.

The deformation in a twisted rod gives rise to elastic forces which counterbalance the applied forces. Since the elements of the rod can turn about its axis, the equilibrium condition is, as we know from mechanics, that the elastic and applied torques are equal. Hence it follows that the magnitude of the torsional deformation must be determined by the applied torque about the rod axis (also called the *torsional torque*). For small deformations (when the angle of shear β is small), Hooke's law is valid and the angle of twist of the rod is proportional to the torsional torque.

The relation between the angle of twist and the torsional torque can be used to measure the latter. This method of measuring torques is widely used in physics in what is called a *torsion balance*. Here the “rod” usually consists of a fine quartz thread of thickness from 1 to 100 μ , having high sensitivity and strength; the angle of twist of the thread is measured from the movement of a light spot reflected from a mirror fixed to the thread. Extremely small torques can be measured by means of such a balance. An intrinsic limit of sensitivity is imposed only by the spontaneous random vibrations of the balance due to unavoidable thermal fluctuations (similar to Brownian motion). As an example, the amplitude of the fluctuation torsional oscillations of a balance with a quartz thread 10 cm long and 1 μ thick is only a fraction of a minute of arc at room temperature.

§104. Plasticity

There is a fundamental difference between compression (or extension) and shear deformations, which may be explained as follows. Let us consider a body undergoing shear, for example a cube of some material placed in a rigid container in the form of an oblique parallelepiped of equal volume. As a result of the shear, the body will contain some stored elastic energy.

It is easily seen that the configuration of the atoms in the deformed cube is not energetically advantageous. In other words, their configuration does not correspond to stable equilibrium (for a given shape of the body). For let us imagine that the container is filled with the material of the cube in molten form. By allowing this to solidify we obtain a body for which the shape of the container is natural and the shape of the cube is unnatural. The new configuration of the atoms is evidently one of lower energy, since it does not possess the shear energy.

We see that a shear deformation is essentially unstable, since the atoms can be arranged within the boundaries of the deformed body in such a way that the energy of the body is decreased.

This conclusion clearly applies only to shear and not to uniform compression. Under compression, the elastic energy results from the change in volume of the body, and therefore can not be eliminated by any movement of the atoms within a fixed volume.

If a shear deformation of a body were to be accompanied by a change in the configuration of atoms such as to eliminate the

elastic energy, then the body would retain its different shape when the external forces were removed, and would not revert to its original shape. Such deformations which remain when the external forces cease to act are called *plastic deformations*.

It is found that plastic deformations do not occur for stresses below a certain value, and the deformation disappears when the external forces cease to act. Such deformations are said to be *elastic*, and the whole of the discussion in §§101–3 refers only to these.

The value of the stress for any given body above which a plastic deformation occurs is called the *elastic limit*. For smaller stresses, the body returns to its original state when the load is removed; for larger stresses, residual plastic deformations remain in the body after removal of the load.

The value of the elastic limit depends not only on the substance but also to a considerable extent on the mode of preparation of the sample, its previous treatment, the presence of impurities etc. For example, the elastic limit of single crystals of aluminium is only about 4 kgf/cm², but that of commercial aluminium is 1000 kgf/cm². The elastic limit of heat-treated carbon steel reaches 6500 kgf/cm².

The elastic limit is very small in comparison with the shear modulus, and the limiting value of the deformation beyond which plasticity occurs is in general very small. For example, the shear modulus of aluminium is 2.5×10^5 kgf/cm². This means, for example, that single crystals of aluminium are elastic only up to relative deformations $\lambda = 4/(2.5 \times 10^5) \sim 10^{-5}$. Steel is elastic up to $\lambda \sim 10^{-2}$.

Plastic deformation itself affects the elastic limit of a body: when a body undergoes a plastic deformation, its elastic limit is raised. This is called *hardening*. For example, the elastic limit of a single crystal of zinc is so small that it can easily be bent with the fingers, but it is not so easily straightened again, since the bending increases the elastic limit. The phenomenon of hardening is, in particular, the basis of the change in properties of a metal by the process of cold working, which consists in plastically deforming it in some way.

Owing to hardening, a body subject to stresses which exceed the elastic limit does not break. It undergoes a plastic deformation which increases until the resulting changes cause the elastic

limit to become equal to the stresses caused by the external forces. We may say that the elastic limit is equal to the stress which caused the last preceding plastic deformation of the body.

Figure 123 shows a diagram of the relation between the stresses p acting in the body and the magnitude λ of the deformation.

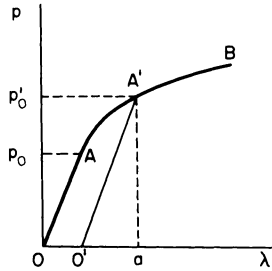


FIG. 123.

If the stress is less than the elastic limit p_0 , the deformation is elastic and obeys Hooke's law (more or less), i.e. λ is proportional to p . This relation is shown in the diagram by the straight line OA .

When the stress becomes greater than p_0 , a plastic deformation of the body occurs, and as the stress increases the relation between λ and p is as shown by the curve AB . Let us suppose that, having reached a point A' on this curve, we then decrease p . The value of $p = p'_0$ corresponding to A' is also the elastic limit acquired by the body through hardening as the load is increased. Thus, when p decreases, there will be no further plastic deformation, and the variation of λ is shown by the straight line $A'O'$, which is parallel to the elastic part AO of the line OB . When the stress becomes zero, there remains some deformation λ_{pl} , which is a plastic deformation. The total deformation at the point A' can be written as the sum of plastic and elastic parts, $\lambda_{pl} = OO'$ and $\lambda_{el} = O'a$.

If the stress is again increased, the same straight line $O'A'$ is traversed until the value p'_0 is reached. On passing the threshold p'_0 we move from the line $O'A'$ to curve $A'B$ and the plastic deformation is increased, the elastic limit being thereby further raised.

With increasing plastic deformation the elastic limit does not, however, increase indefinitely: there is a maximum value of the

elastic limit which cannot be exceeded. This is called the *yield point*. Under a stress of this amount, the deformation of the body increases continuously and it begins to flow like a liquid. By applying high pressures it is possible, for example, to cause a metal to flow in a jet from an aperture in the cylinder of a hydraulic press.

It is clear that stresses exceeding the yield point can never occur in a body for any deformation (except, of course, in uniform compression).

The yield point may sometimes not be reached, since the body may fracture much sooner. In order to observe the yield phenomenon, it is best to use such deformations as unilateral compression or torsion. Simple stretching, on the other hand, easily causes fracture.

The presence of small, frequently microscopic, cracks in a body plays an important part in fracture. These cracks may be either on the surface of the body or within it, for example slight gaps between the grains of a polycrystalline body. Such cracks act as levers to cause a considerable concentration of the external forces applied to the body: it is comparatively easy for the elastic stresses at the sharp end of a crack to reach values sufficient for further rupture of atomic bonds and lengthening of the crack, ultimately leading to complete fracture of the body. The importance of the state of the surface of the body with regard to fracture is clearly shown by an experiment with a rock-salt crystal: if the crystal is immersed in water, the salt dissolves from its surface, the cracks present on the surface are eliminated, and the crystal under water is considerably more difficult to break than a crystal in air.

Plastic deformation near the ends of cracks may blunt their points and thus decrease to some extent the concentration of elastic stresses near them. In this sense plasticity aids the resistance of a body to fracture, as is shown by the temperature dependence of brittleness in metals. For example, steel, which is difficult to break at ordinary temperatures, becomes brittle at low temperatures. This effect is largely due to decreasing plasticity at low temperatures, which will be further discussed in §106.

§105. Defects in crystals

The very fact that the plastic properties of a body depend considerably on its previous treatment, the presence of impurities, etc., indicates that these properties are closely related to features

of the crystal structure of actual bodies which distinguish them from ideal bodies.

Departures from the ideal crystal structure are called *defects*. The simplest type, which may be called point defects, consist in the absence of an atom from a lattice point (a free *vacancy*) or the replacement of the “correct” atom at a lattice point by a different (impurity) atom, the entry of an extra atom between lattice points, and so on. The departure from the regular structure of the lattice extends over a distance of the order of several lattice periods around such a point.

The most important defects as regards the mechanical properties of solids are, however, of another kind, which may be called line defects, since the departure from the regular structure of the lattice is concentrated near certain lines. These are *dislocations*.

The dislocation shown in Fig. 124 may be regarded as a lattice defect caused by the presence in the lattice of an extra crystal half-plane inserted between two “regular” planes (layers of

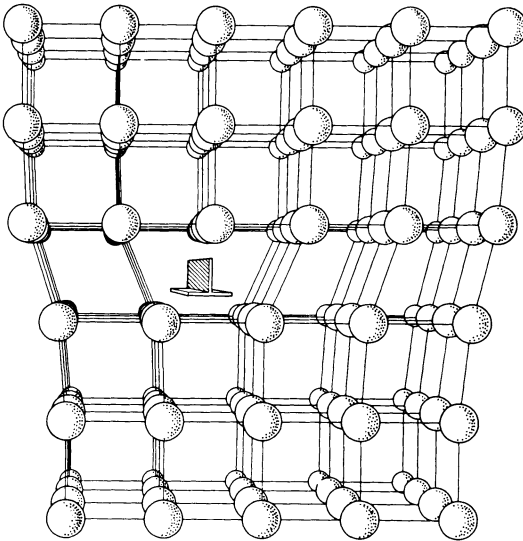


FIG. 124.

atoms). The line of the dislocation (which in this case is called an *edge dislocation*) is a straight line perpendicular to the plane of the diagram, shown by the symbol \perp ; the “extra” layer of atoms lies above this symbol. The dislocation may also be

regarded as the result of displacing the upper part of the crystal shown diagrammatically in Fig. 125a by one lattice period (Fig. 125b).

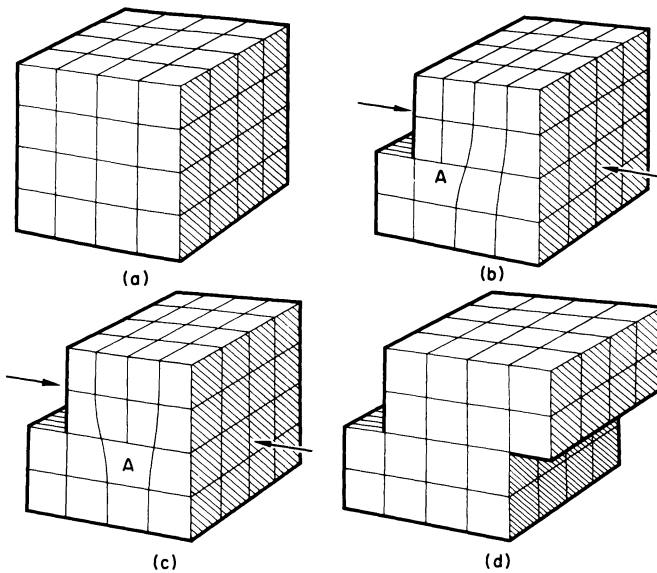


FIG. 125.

Another type of dislocation may be visualised as the result of “cutting” the lattice along a half-plane and then displacing the parts of the lattice on either side of the cut by one lattice period along the edge of the cut (Fig. 126). The edge of the cut is then called a *screw dislocation* and is shown by the broken line in Fig. 126. The presence of such a dislocation converts

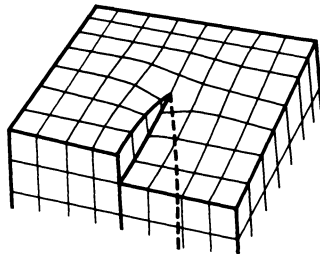


FIG. 126.

the crystal plane in the lattice into a helicoidal surface, like a spiral staircase without the steps.

In an edge dislocation the displacement is perpendicular to the dislocation line, but in a screw dislocation it is parallel to this line. Any intermediate case between these two extremes is possible. The dislocation lines need not be straight: they may be curves or even closed loops.

There are various methods of directly observing dislocations. For example, in transparent crystals this can be done by creating supersaturated solid solutions of certain substances. The impurity atoms tend to be deposited as colloidal particles, which grow mainly at the places where the basic lattice structure is perturbed; thus the colloidal particles of impurities are concentrated along dislocation lines and render them visible. Another method is based on the etching of the crystal surface by suitable reagents. The surface is more easily attacked at points where the crystal structure is perturbed. This leads to the formation of visible pits at points where the dislocation lines reach the surface of the crystal.

Screw dislocations often play a decisive part in the process of growth of crystals from a liquid or a supersaturated vapour.

It has been shown in §99 how the formation of a new phase within the original phase must begin with nucleation. A similar situation must occur in the growth of a crystal. The formation of a new layer of atoms on a perfectly regular crystal surface cannot begin simply with the deposition of individual atoms on the surface: such atoms, having neighbours on one side only, would be under conditions which would be energetically very unfavourable, and would not remain on the surface. A stable "nucleus" for a new layer of atoms on the surface of the crystal must contain immediately a sufficient number of atoms, and the chance occurrence of such nuclei may be comparatively rare. If, however, the edge of a screw dislocation appears on the surface of the crystal, it provides a ready-made step (one atomic layer in height) to which new atoms can easily attach themselves, and no nuclei are therefore necessary. The rate of attachment of new atoms is approximately the same all along the edge of the step. The crystal therefore grows spirally, as shown diagrammatically in Figs. 127a–d. At any time there is a free step on the surface of the crystal, which can therefore grow without limit. The rate

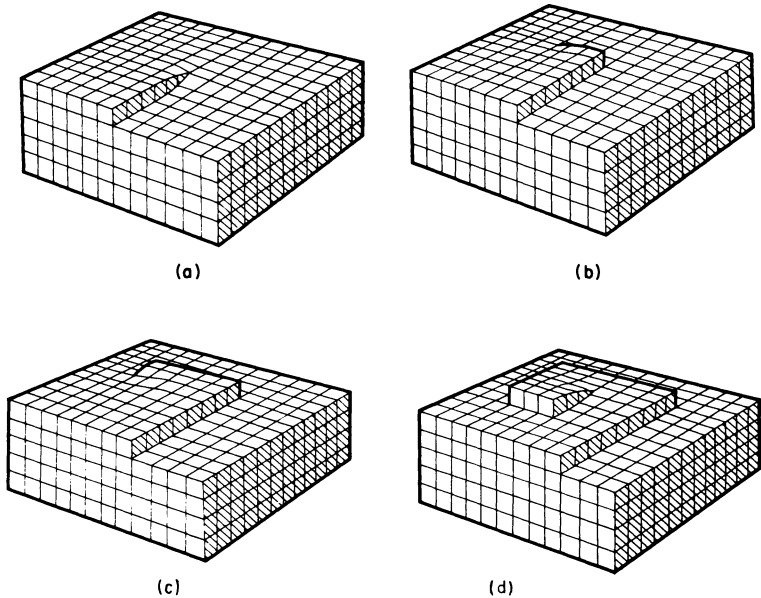


FIG. 127.

of growth is very much higher than that of a process depending on nucleation.

§106. The nature of plasticity

Groups of parallel lines can often be observed on the surface of a single crystal undergoing a plastic shear deformation. These lines are the traces of the intersection of the surface of the body with the *slip planes* along which some parts of the crystal slide as a whole relative to other adjoining parts. Thus the plastic deformation is non-uniform: large displacements in shear occur only along planes at a comparatively large distance apart, while the parts of the crystal which lie between these planes undergo almost no deformation. Fig. 128 is a diagram of the deformation of a body as a result of slipping of this kind.

The configuration of the slip planes is closely related to the structure of the crystal lattice. In any crystal, slipping occurs almost entirely along certain planes; for example, in the NaCl crystal these are (110) planes, while in metal crystals with face-centred cubic lattices they are (111) planes.

What is the mechanism whereby one part of a crystal slips relative to another? If this were to take place simultaneously over the whole slip plane, very large stresses would be necessary.

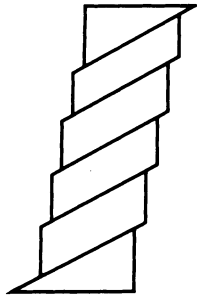


FIG. 128.

The change from one equilibrium configuration of atoms to another (say, from that shown in Fig. 125a to that in Fig. 125d) would have to occur by means of a large elastic deformation in which the relative displacements (in the region near the slip plane) would reach values $\lambda \sim 1$. This would require stresses of the order of the shear modulus G .

In reality, the elastic limits of actual bodies are usually 10^2 to 10^4 times less than their shear moduli, and so relatively small stresses are necessary in effecting a shear. This is possible because slip actually takes place by the movement of dislocations in crystals.

The simplest form of this mechanism is indicated by the sequence in Figs. 125 a-d. If the crystal contains an edge dislocation (passing through a point A and at right angles to the front face of the crystal), the movement of this dislocation in the slip plane from left to right through the body causes a displacement of the upper part of the crystal relative to the lower part by one lattice period. The movement of the dislocation involves only a relatively slight reconstruction of the lattice, which affects only the atoms near a single line. This process may be compared to the movement of a wrinkle in a carpet: the wrinkle moves more easily than the whole carpet, but the effect of moving the wrinkle from one end of the carpet to the other is to shift the whole carpet a certain distance.

Thus the plasticity of a solid depends on the presence of dislocations in it and on the possibility of their free movement.

This movement may, however, be retarded by various obstacles, such as impurity atoms dissolved in the lattice or small solid inclusions in the body. Dislocations are also slowed down by intersecting one another, and by the grain boundaries in a polycrystalline body. At the same time, the interaction of dislocations with one another and with other defects gives rise to new dislocations. These processes are very important, since they support the development of a plastic deformation; otherwise, the deformation would cease as soon as all the dislocations existing in the body had been "utilised".

The number of dislocations in a body is described by the dislocation density, which is the number of dislocation lines intersecting a unit area within the body. This number varies widely, from $10^2 - 10^3 \text{ cm}^{-2}$ in the most perfect pure single crystals to $10^{11} - 10^{12} \text{ cm}^{-2}$ in heavily deformed (cold-worked) metals.

It is clear from the above discussion that pure single crystals will have the lowest strength (i.e. the lowest elastic limit), since the dislocation density in them is comparatively low, and so there is practically no interference between the dislocations in their motion. Hardening of the material can be achieved by dissolving impurities in it or depositing microscopic solid inclusions, or by reducing the grain size. For example, the strength of iron is increased (in various kinds of steel) by dissolving in it carbon atoms or microscopic inclusions of iron carbide which are deposited in the process of solidification.

Plastic deformation itself damages the crystal lattice, increasing the number of defects in crystals and thereby impeding the further movement of dislocations. This is the reason for the phenomenon of hardening by deformation, including the hardening of metals by cold working (*work-hardening*).

The hardening achieved by plastic deformation is not maintained for an indefinite time, however. The most stable state of a body is the undisturbed ideal crystal, which is the state having the least energy. Thus perturbed crystals exhibit what is called *recrystallisation*. The structural defects are "healed" and the large grains in a polycrystalline body increase in size at the expense of the smaller ones, resulting in a system which is less defective and therefore of lower strength. Recrystallisation occurs more rapidly at high temperatures, and especially rapidly

at temperatures fairly near the melting point (for example, in the annealing of metals). At low temperatures there is practically no recrystallisation. The effect of recrystallisation is gradually to eliminate the hardening, and if the body is subject to a steady load it will slowly flow.

The temperature also has a marked effect on the movement of dislocations. Since this movement involves the overcoming of potential barriers by the atoms (which change their configuration near the moving dislocation line), it is a process of the activation type (cf. §91), and therefore is rapidly stopped by lowering the temperature, thus decreasing the plasticity of the body.

The methods described above for increasing the strength of a material are based on the creation of obstacles to the movement of dislocations. The opposite means of hardening is also possible, namely to produce a single crystal which contains no dislocations at all. Such a crystal should in principle have the maximum possible elastic limit: its plastic deformation could be brought about only by simultaneous slipping along entire planes, which, as already mentioned, would require the application of extremely large stresses.

This ideal state is approached by what are called *whiskers*. These are extremely thin thread-like crystals with thicknesses of the order of microns. They are formed both by metals and by non-metals, and can be obtained in various ways: by precipitation of slightly supersaturated vapours of pure metals at appropriate temperatures in an inert gas medium, by slow precipitation of salts from solutions, and so on. In many cases such crystals appear to grow round individual screw dislocations in the manner described in §105. The dislocation along the axis of the whisker does not affect its mechanical properties when it is stretched, and the crystal behaves practically as an ideal one.

It is clear from the above discussion that all these plasticity properties relate only to crystalline bodies. Amorphous bodies, such as glass, are not able to undergo plastic deformation, and are said to be *brittle*. Their inelastic behaviour consists of either fracture or a slow flow under the prolonged action of forces, in accordance with the fact that amorphous bodies are actually liquids of very high viscosity.

§107. Friction of solids

The sliding of a solid body on the surface of another body is always accompanied by the conversion of its kinetic energy into heat, and in consequence the motion is gradually retarded. This phenomenon can be described from the purely mechanical point of view as being due to a certain force which impedes the motion, called a *frictional force*. Physically, friction is the result of complex processes which occur on surfaces which rub together.

Experiment shows that the friction between solid bodies usually obeys certain simple laws. It is found that the total frictional force F_{fr} acting between moving bodies is proportional to the force N which presses the bodies together, and does not depend on the area of contact between the bodies or on the speed of the motion:

$$F_{\text{fr}} = \mu N.$$

The quantity μ is called the *coefficient of friction*; it depends only on the properties of the surfaces which rub together. This relation is usually satisfied to a good approximation over a wide range of experimental conditions (loads and rates of sliding), but deviations from it are also found.

Friction depends considerably on the way in which the rubbing surfaces have been treated and on their present state (whether contaminated, and by what). For example, the coefficient of friction between metal surfaces is usually in the range from 0.5 to 1.5. These values, however, are for metal surfaces exposed to the air. Such surfaces are always contaminated by oxides, adsorbed gases, etc., which impair the conditions of contact. Experiment shows that completely clean metal surfaces prepared by heating in vacuum show very high friction in sliding, and sometimes “stick” completely.

There is probably no single universal mechanism of friction, and the nature of friction is different for surfaces of different types and with different previous treatment. As an illustration we shall describe the mechanism of friction for certain metals.

Experiment shows that metal surfaces always exhibit irregularities which are large in comparison with molecular distances. Even for surfaces prepared and polished in the best possible way, the depth of the irregularities is 100–1000 Å, and rubbing surfaces in engineering usually have much greater non-uniformities.

When bodies touch, the actual contact between them occurs only at the “peaks” of these non-uniformities. Thus the area of actual contact S_0 may be very small in comparison with the total nominal area of contact S ; the ratio may be 10^{-4} or 10^{-5} . In plastic metals, even under small loads, the “peaks” of the non-uniformities are deformed and flattened until the true pressure acting on them decreases to a certain value p_{lim} below which the deformation ceases. The area of contact S_0 is determined by the condition $p_{\text{lim}}S_0 = N$, and is therefore proportional to the load N . In the regions of actual contact, the forces of molecular cohesion bring about a strong “adhesion” of the bodies. During sliding there is a continual separation and formation of fresh regions of contacts. The force required to break contact is proportional to the area of contact S_0 , and therefore to the load N .

The frictional force during motion must be distinguished from the force needed at the beginning of the motion in order to start the body from rest. This *limiting friction* is also proportional to the load, but the coefficient is somewhat greater than in motion (although the difference is not more than 10–20%).

It should be emphasised that the whole of the above discussion refers to friction between *dry* surfaces of solid bodies. It bears no relation to the friction between lubricated surfaces separated by a layer of fluid. In the latter case the frictional force is due to the viscosity of the liquid; a simple example of this type of friction will be discussed in §119.

As well as sliding friction, there is also the friction that occurs when one body rolls on another. Let us consider a cylinder of radius r rolling on a plane. In order to overcome the frictional force and maintain steady rolling, a force F must be applied, which is described by the torque K about the instantaneous line of contact between the cylinder and the plane; if the force is applied to the axis of the cylinder, then $K = rF$. The torque K is a measure of the rolling friction; it is found to be proportional to the force N which presses the rolling body to the surface on which it rolls:

$$K = \gamma N.$$

The coefficient γ depends on the two bodies in contact; it clearly has the dimensions of length.

DIFFUSION AND THERMAL
CONDUCTION

§108. The diffusion coefficient

In the preceding chapters we have discussed mainly the properties of bodies in thermal equilibrium. This chapter and the next deal with processes by means of which a state of equilibrium is reached, called *kinetic processes*. These are all essentially irreversible processes, since they bring a body closer to equilibrium.

If a solution has different concentrations at different points, the thermal motion of the molecules causes mixing of the solution in the course of time: the solute moves from regions of higher to regions of lower concentration, until the composition of the solution becomes uniform throughout its volume. This process is called *diffusion*.

For simplicity, let us assume that the concentration of the solution (denoted by c) varies only in one direction, which we shall take as that of the x axis. The *diffusion flux* j is defined as the quantity of solute passing per unit time through a surface of unit area perpendicular to the x axis, and will be taken as positive if the flux is in the positive direction of this axis, and negative if it is the opposite direction. Since matter passes from regions of higher to regions of lower concentration, the sign of the flux is opposite to that of the derivative dc/dx (called the *concentration gradient*): if the concentration increases from left to right, the flux is to the left, and conversely. If $dc/dx = 0$, i.e. the concentration of the solution is constant, there is no diffusion flux.

All these properties are included in the following relation between the diffusion flux and the concentration gradient:

$$j = -D \, dc/dx.$$

Here D is a constant coefficient called the *diffusion coefficient*. This relation describes the properties of diffusion “phenomenologically”, that is, from its external manifestations. We shall see below (§113) how a similar expression for the flux can be derived directly by considering the molecular mechanism of diffusion.

The flux j in the above formula may be defined in any manner: as the mass of solute passing through unit area, as the number of solute molecules, and so on, but the concentration c must then be defined in a similar manner as the mass or number of molecules of solute per unit volume. Then it is evident that the diffusion coefficient will not depend on the way in which the flux and the concentration are defined.

The dimensions of the diffusion coefficient may be found as follows. Let j be the number of solute molecules passing through unit area per unit time. Then $[j] = 1/\text{cm}^2\cdot\text{sec}$. The concentration is the number of solute molecules per unit volume, with dimensions $[c] = 1/\text{cm}^3$. Comparing dimensions on the two sides of the equation $j = -D \, dc/dx$, we find

$$[D] = \text{cm}^2/\text{sec}.$$

When speaking of diffusion, we imply that it occurs in a medium at rest, so that the equalising of the concentration occurs only because of the random thermal motion of the individual molecules. It is assumed that the liquid (or gas) is not mixed by any external interaction which causes it to move.

Such mixing may occur in a liquid, however, because of gravity. If a light liquid such as alcohol is carefully poured on water, the liquids will mix by diffusion, but if water is poured on alcohol, streams of water (the heavier liquid) will descend and streams of alcohol will rise.

Thus gravity may cause the composition of a medium to be equalised by movement. This is called *convection*; it equalises the concentration much more rapidly than diffusion.

§109. The thermal conductivity

The process of *thermal conduction* is akin to diffusion. If the temperature is different at different points in a body, a heat flux occurs from hotter to colder regions, and continues until the

temperature is the same throughout the body. Here again the mechanism of the process is based on the random thermal motion of the molecules: molecules belonging to the hotter parts of the body collide with molecules in adjoining colder parts and transmit to them part of their energy.

As in the discussion of diffusion, it is assumed that thermal conduction takes place in a medium at rest. In particular, it is assumed that the medium contains no pressure variations which would cause motion in it.

Let us suppose that the temperature T of the medium varies only in one direction, which we again take as that of the x axis. The heat flux q is defined as the quantity of heat passing per unit time through unit area perpendicular to the x axis. Just as for diffusion, the relation between the heat flux and the temperature gradient dT/dx is

$$q = -\kappa dT/dx.$$

Here again the minus sign appears because the direction of the heat flux is opposite to that in which the temperature increases: heat flows in the direction of decreasing temperature. The coefficient κ is called the *thermal conductivity*.

If the quantity of heat is measured in ergs, the heat flux will be measured in erg/cm².sec, and the dimensions of the thermal conductivity are therefore

$$\begin{aligned} [\kappa] &= [\text{erg/cm}\cdot\text{sec}\cdot\text{deg}] \\ &= [\text{g}\cdot\text{cm}/\text{sec}^3\cdot\text{deg}]. \end{aligned}$$

The thermal conductivity determines the rate of flow of heat from hotter to colder regions. The change in temperature of a body is equal to the quantity of heat gained, divided by the specific heat. Thus the rate of equalisation of the temperature at different points in the body is governed by the thermal conductivity divided by the specific heat per unit volume, i.e. the quantity

$$\chi = \kappa/\rho c_p,$$

where ρ is the density and c_p the specific heat per unit mass (at constant pressure, since thermal conduction at constant pressure is being discussed). This quantity is called the *thermal diffusivity*.

It is easily seen to have the dimensions

$$[\chi] = \text{cm}^2/\text{sec},$$

which are the same as those of the diffusion coefficient. This is natural, since if both sides of the relation $q = -\kappa dT/dx$ are divided by ρc_p , the ratio $q/\rho c_p$ on the left-hand side may be regarded as a "temperature flux", i.e. the flux of the quantity whose gradient appears on the right. Thus the coefficient χ is a kind of diffusion coefficient for temperature.

As with diffusion, the action of gravity may cause convective mixing of a non-uniformly heated liquid (or gas). This occurs when the liquid is heated below (or cooled above): the hotter and therefore less dense lower layers of the liquid rise and are replaced by descending currents of colder liquid. The equalisation of temperature by convection occurs, of course, much more rapidly than by thermal conduction.

As examples, the following table shows the values of the thermal conductivity for a number of liquids and solids (at room temperature). These values are given in units of J/cm. sec. deg, i.e. the heat flux is defined as the energy in joules transported through 1 cm² in 1 sec.

Water	6.0×10^{-3}	Lead	0.35
Benzene	1.5×10^{-3}	Iron	0.75
Glass	$4 \text{ to } 8 \times 10^{-3}$	Copper	3.8
Ebonite	1.7×10^{-3}	Silver	4.2

The very high thermal conductivity of metals should be noted. The reason for this is that in metals, unlike other bodies, heat is transferred by the thermal motion of free electrons, and not of atoms. The effectiveness of heat conduction by electrons is due to their high velocity, of the order of 10^8 cm/sec, which is much higher than the ordinary thermal velocities of atoms and molecules (10^4 to 10^5 cm/sec).

§110. Thermal resistance

The simple relation given above between the heat flux and the temperature gradient makes possible the solution of various problems relating to thermal conduction.

Let us consider a layer of material (of thickness d) between two parallel planes, each of area S , and assume that these boundary planes are maintained at different temperatures T_1 and T_2 (with $T_1 > T_2$). The thermal conductivity of the substance is in general a function of temperature, but we shall suppose that the difference between the temperatures T_1 and T_2 is not very great, so that we may neglect the variation in the conductivity across the thickness of the layer and regard κ as a constant.

Let the x axis be taken across the thickness of the layer, and let x be measured from the plane at temperature T_1 . It is evident that a temperature distribution depending only on x will be established in the layer of material, and a heat flux through the layer from T_1 to T_2 will exist. Let us find the relation between this flux and the temperature difference $T_1 - T_2$ which causes it.

The total heat flux Q through the whole cross-section of the layer (parallel to the boundary planes) per unit time is equal to the product qS of the flux q per unit area and the total area S of the cross-section. Using the relation between q and the temperature gradient, we can write

$$Q = -\kappa S \, dT/dx.$$

The flux Q is clearly independent of x , since no heat is absorbed in passing through the layer and none is evolved within the layer; the total quantity of heat passing per unit time through any surface which intersects the entire layer must therefore be the same. From the above equation we therefore have

$$T = -(Q/\kappa S)x + \text{constant},$$

i.e. the temperature varies linearly across the thickness of the layer. When $x = 0$, i.e. on one of the boundary planes, we must have $T = T_1$; hence the constant is equal to T_1 , and

$$T = T_1 - (Q/\kappa S)x.$$

At the other boundary plane ($x = d$) we must have $T = T_2$, i.e.

$$T_2 = T_1 - (Q/\kappa S)d.$$

Hence

$$Q = (\kappa S/d)(T_1 - T_2).$$

This formula gives the required relation between the heat flux Q and the temperature difference across the layer.

Let us now consider a layer of material bounded by two concentric spheres (of radii r_1 and r_2) maintained at temperatures T_1 and T_2 . Figure 129 shows a central cross-section. The temperature at any point within the layer is evidently a function only of the distance r from the centre of the spheres.

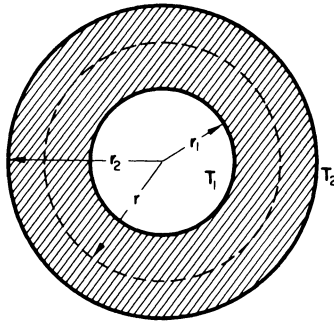


FIG. 129.

Since the only coordinate on which the temperature depends in this case is r , the heat flux q is everywhere in the radial direction, and is

$$q = -\kappa dT/dr.$$

The total heat flux through a spherical surface of radius r concentric with both spheres and lying between them is

$$Q = 4\pi r^2 q = -4\pi \kappa r^2 dT/dr,$$

whence

$$dT/dr = -Q/4\pi \kappa r^2.$$

As in the previous case, the total heat flux through any closed surface enclosing the inner sphere must be the same, and Q is therefore independent of r . The above equation then gives

$$T = \frac{Q}{4\pi\kappa r} + \text{constant.}$$

The constant is determined by the condition that $T = T_1$ for $r = r_1$, so that

$$T = T_1 + \frac{Q}{4\pi\kappa} \left(\frac{1}{r} - \frac{1}{r_1} \right).$$

Finally, from the condition that $T = T_2$ for $r = r_2$ we obtain the following relation between the total heat flux and the temperature difference across the layer:

$$Q = \frac{(T_1 - T_2) \cdot 4\pi\kappa}{1/r_1 - 1/r_2}.$$

In particular, if $r_2 = \infty$, i.e. if there is an infinite medium round a spherical surface of radius r_1 (T_2 in this case being the temperature at infinity), the expression for the heat flux becomes

$$Q = 4\pi\kappa r_1 (T_1 - T_2).$$

The ratio of the temperature difference at the boundaries of a body to the total heat flux is called the *thermal resistance* of the body. The above formulae show that the thermal resistance of a plane slab is $d/\kappa S$, and that of a spherical layer is

$$\frac{1}{4\pi\kappa} \left(\frac{1}{r_1} - \frac{1}{r_2} \right).$$

Entirely similar results are evidently obtained for diffusion in a solution bounded by two planes or two spherical surfaces on which given concentrations are maintained. In the above formulae we need only replace the temperature by the concentration, the heat flux by the diffusion flux and κ by the diffusion coefficient D .

Let us apply these formulae to the problem of rate of melting, and consider a piece of ice immersed in water at a temperature T_1 above 0°C . Since equilibrium between ice and water is possible (at atmospheric pressure) only at a definite temperature $T_0 = 0^\circ\text{C}$, the water immediately adjoining the ice will be at this temperature. At increasing distances from the ice, the water temperature is greater and tends to T_1 . There will be a heat flux from the water to the ice. On reaching the ice, the heat is absorbed as the heat of fusion necessary to convert ice into water. For example, if the piece of ice is spherical (with radius r_0), it will receive per unit time from the surrounding water (which we regard as an infinite medium) a quantity of heat

$$Q = 4\pi\kappa r_0(T_1 - T_0).$$

Dividing this by the heat of fusion, we find the quantity of ice which melts per unit time. Thus the rate of melting is determined by the process of thermal conduction in the surrounding water.

Similarly, the rate of dissolution of a solid in a liquid is determined by the rate of diffusion of solute in the liquid. Near the surface of the solid, a thin layer of saturated solution is immediately formed; further dissolution takes place as the solute diffuses from this layer into the surrounding liquid. For example, if the solid is a sphere of radius r_0 , the total diffusion flux J from the sphere into the solvent, which is the quantity of substance dissolving per unit time, is

$$J = 4\pi D r_0 c_0.$$

Here c_0 is the concentration of the saturated solution, and the concentration in the liquid at a great distance from the sphere is taken to be zero.

Processes of diffusion and thermal conduction also determine the rate of evaporation of a liquid drop in a gas of another substance, such as air. The drop is surrounded by a layer of saturated vapour, from which the substance slowly diffuses into the surrounding air. The process of heat transfer from the air to the drop is also of importance.

These examples are typical in that the rates of phase transitions occurring under steady-state conditions are usually determined by processes of diffusion and thermal conduction.

§111. The equalisation time

If the concentration of a solution is different at different points, then, as we know, the composition will be equalised in the course of time by diffusion. Let us determine the order of magnitude of the time t required for this process. This may be done from considerations of the dimensions of the quantities on which this time can depend.

First of all, it is evident that the time t cannot depend on the actual concentrations of the solution, for if all the concentrations are changed by a given factor, the diffusion flux which equalises the concentrations is changed by the same factor, and the equalisation time therefore remains unchanged.

The only physical quantities on which the time t of diffusion equalisation can depend are the diffusion coefficient D in the medium concerned and the size of the region in which the concentrations are different; let the linear size of this region be of order of magnitude L .

The dimensions are $[D] = \text{cm}^2/\text{sec}$, $[L] = \text{cm}$. It is evident that only one combination having the dimensions of time can be formed from these quantities, namely L^2/D , and this must give the order of magnitude of the time t :

$$t \sim L^2/D.$$

Thus the time for equalisation of concentrations in a region of size L is proportional to the square of L and inversely proportional to the diffusion coefficient.

This question can be inversely stated as follows. Let us suppose that at some initial instant there is a certain quantity of solute concentrated in a small region of the solvent. In time this accumulation of solute will be dispersed by the effect of diffusion, and will be distributed throughout the whole large volume of the solvent. What is the mean distance L traversed by the diffusing substance in a time t ? That is, we now wish to find the distance from the time, not the time from the distance. The answer is clearly given by the same formula, which must now be written

$$L \sim \sqrt{(Dt)}.$$

Thus in a time t the diffusing substance spreads to a distance proportional to \sqrt{t} .

This relation may also be regarded in another way. Let us consider any one molecule of solute in the solution. Like all molecules it has a random thermal motion. We may ask what is the order of magnitude of the distance which this molecule can traverse from its initial position in time t ; in other words, what is the mean straight-line distance between the initial and final positions of a molecule which has moved for a time t . Instead of considering a single molecule, let us suppose that there is a very large number of molecules close together. Then, as we have seen, in the course of time these molecules will move apart in all directions by diffusion, and the average distance travelled is $L \sim \sqrt{Dt}$. This distance L is clearly also the mean distance that each molecule moves from its original position in time t .

This result applies not only to molecules of solute but also to any particles suspended in a liquid and executing Brownian motion.

The above discussion has referred entirely to diffusion, but the same arguments apply also to thermal conduction. We have seen in §109 that in the propagation of heat the diffusion coefficient is replaced by the thermal diffusivity χ . Thus the temperature equalisation time in a body of linear size L is

$$t \sim L^2/\chi \sim L^2\rho c_p/\kappa.$$

This relation also can be inverted as was done above for the case of diffusion. In this connection let us consider the following problem. We assume that fluctuations of temperature with some frequency ω are artificially created on the surface of a body. These fluctuations will penetrate into the body, producing what is called a *thermal wave*. The amplitude of the fluctuations, however, will be damped with increasing depth in the body, and the question is to what depth L the fluctuations penetrate. Here the characteristic time is the period of the fluctuations, i.e. the reciprocal of the frequency. Substituting $1/\omega$ for t in the relation between the distance of heat propagation and the time, we obtain

$$L \sim \sqrt{\chi/\omega}.$$

This is the solution to the problem.

§112. The mean free path

Turning now to discuss thermal conduction and diffusion in gases, we must first consider the nature of the interaction between gas molecules in somewhat more detail than hitherto.

Gas molecules interact by means of collisions. During the greater part of the time, the molecules are comparatively far apart and move as if free, scarcely interacting at all. The molecules interact only during short intervals of time when they collide with one another. In this respect a gas differs from a liquid, in which the molecules are continuously interacting, and they cannot be said to undergo separate "collisions".

Molecules may collide in various ways. Strictly speaking, in each passage of molecules at not too great a distance they undergo some change in velocity, and the concept of a "collision" is therefore not entirely precise. In order to make the concept more definite, we shall regard as collisions only those cases where the molecules pass so close that the interaction considerably alters their motion, i.e. their velocities are considerably changed in magnitude or direction.

Collisions between molecules in a gas occur completely randomly, and the distance travelled by a molecule between two successive collisions may therefore have any value. We can, however, define a mean value of this distance, which is called the *mean free path* of the molecules, and is an important molecular-kinetic property of the gas; it will be denoted by l . As well as the mean free path, we may consider also the mean time τ between two successive collisions. In order of magnitude, evidently,

$$\tau \sim l/v,$$

where v is the mean velocity of thermal motion of molecules.

Let us consider two colliding molecules, regarding one of them as being at rest in a certain plane, and the other as crossing this plane. As explained above, the molecules will be said to collide only when they pass so close that their motion is considerably altered. This means that the moving molecule collides with the stationary one only if it meets the plane somewhere within a certain small region around the fixed molecule. This "target" area which the molecule must strike is called the *effective*

cross-section (or simply the *cross-section*) for collisions, and will be denoted by σ .

As an example, let us determine the collision cross-section for molecules regarded as solid spheres of radius r_0 . The greatest distance between the centres of two spheres at which they can pass and still touch is $2r_0$. Thus the "target" area which the molecule must strike if a collision is to occur is a circle of radius $2r_0$ round the centre of the stationary molecule. Thus the collision cross-section in this case is

$$\sigma = 4\pi r_0^2,$$

or four times the cross-sectional area of the sphere.

In reality, of course, molecules are not solid spheres, but since the interaction force between two molecules decreases very rapidly with increasing distance between them, collisions occur only if the molecules almost "graze" each other. The collision cross-section is therefore of the order of magnitude of the cross-sectional area of the molecule.

Let a molecule traverse a distance of unit length in its motion, and let us imagine the molecule as sweeping out a volume of unit length and cross-sectional area σ ; the magnitude of this volume is also σ . The molecule collides with all molecules lying within this cylinder. Let n be the number of molecules per unit volume. Then the number of molecules in the volume σ is $n\sigma$, and the molecule therefore undergoes $n\sigma$ collisions per unit length of path. The mean distance between two collisions, i.e. the mean free path, is in order of magnitude

$$l \sim 1/n\sigma.$$

It is seen from this expression that the mean free path is inversely proportional to the gas density and depends on no other quantity.

It must be remembered, however, that this last statement is valid only if the cross-section is assumed constant. Because the repulsion forces increase very rapidly as the molecules approach, molecules usually behave qualitatively as elastic solid particles, which interact only when they "graze" each other. Under these conditions the collision cross-section is in fact a

constant (depending only on the nature of the molecules). There are also, however, weak forces of attraction between molecules at greater distances. As the temperature decreases, the velocities of the gas molecules become less, and thus the duration of a collision between two molecules (passing at a given distance) increases. Because of this "lengthening" of the collision the motion of the molecules may be considerably changed even if they pass relatively far from each other. Thus, when the temperature decreases, the collision cross-section increases somewhat. For example, in nitrogen and oxygen σ increases by about 30% when the temperature falls from $+100^{\circ}\text{C}$ to -100°C , and in hydrogen by 20%.

For air at 0°C and atmospheric pressure, $n \approx 3 \times 10^{19}$. The cross-section $\sigma \approx 5 \times 10^{-15} \text{ cm}^2$, and therefore the mean free path of the molecules $l \approx 10^{-5} \text{ cm}$. The mean thermal velocity of the molecules $v \approx 5 \times 10^4 \text{ cm/sec}$, and the time between collisions is accordingly $\tau \approx 2 \times 10^{-10} \text{ sec}$. The mean free path increases rapidly with decreasing pressure. For instance, at an air pressure of 1 mm Hg $l \approx 10^{-2} \text{ cm}$; in a high vacuum of the order of 10^{-6} mm Hg pressure, the mean free path reaches values of tens of metres.

§113. Diffusion and thermal conduction in gases

By means of the concept of the mean free path we can determine the order of magnitude of the diffusion coefficient and the thermal conductivity and ascertain how they depend on the state of the gas. Let us take first the diffusion coefficient and consider a mixture of two gases whose total pressure is everywhere constant but whose composition varies in one direction, which we take as that of the x axis.

Let n_1 be the number of molecules of one of the gases in the mixture per unit volume; this number is a function of the co-ordinate x . The diffusion flux j is the number of molecules passing per unit time through unit area perpendicular to the x axis and moving in the positive direction of that axis, minus the corresponding number moving in the negative direction.

The number of molecules passing through unit area per unit time is equal in order of magnitude to $n_1 v$, where v is the mean thermal velocity of the molecules. Here we may suppose that the number of molecules crossing this area from left to right

is determined by the density n_1 at the point where the molecules underwent their last collision, i.e. at a distance l to the left of the area; similarly, for molecules going from right to left we must take the value of n_1 at a distance l to the right of the area. If the coordinate of the area itself is x , the diffusion flux is given by

$$j \sim vn_1(x-l) - vn_1(x+l).$$

Since the mean free path l is a small quantity, the difference $n_1(x-l) - n_1(x+l)$ may here be replaced by $-l \, dn_1/dx$. Thus

$$j \sim -vl \, dn_1/dx.$$

Comparison of this expression with the formula $j = -D \, dn_1/dx$ shows that the diffusion coefficient in a gas is in order of magnitude

$$D \sim vl.$$

The mean free path $l \approx 1/n\sigma$, where n is the total number of molecules of the two gases per unit volume. Thus D may also be written as

$$D \sim v/n\sigma.$$

Finally, the equation of state of an ideal gas shows that the number density of molecules in it is $n = p/kT$, so that

$$D \sim vkT/p\sigma.$$

The diffusion coefficient in a gas is therefore inversely proportional to its pressure (at a given temperature). Since the thermal velocity of the molecules is proportional to \sqrt{T} , the diffusion coefficient increases with temperature as $T^{3/2}$ if the collision cross-section may be regarded as constant.

The following comment should be made concerning the foregoing derivation. In calculating j we have argued as if only one gas were present, whereas in reality there is a mixture of two gases. It is therefore, strictly speaking, uncertain to which of the two gases the quantities σ and v pertain. Since only the order of

magnitude of the diffusion coefficient is being estimated, this point is unimportant if the molecules of the two gases are similar in mass and size, but it may become significant if there is a great difference between them. A more detailed discussion shows that in this case v must be taken as the greater of the thermal velocities (i.e. the velocity of the molecules of smaller mass), and σ as the greater of the cross-sections.

As well as mutual diffusion of different gases, there can occur mutual diffusion of different isotopes of the same substance. Since the only difference between the isotopic molecules is the relatively slight difference in mass, this is a type of diffusion of gas molecules in their own gas, called *self-diffusion*. The difference in mass of the molecules here acts in practice only as a "label" whereby one molecule may be distinguished from others.

The self-diffusion coefficient of a gas is given by the same formula

$$D \sim vl,$$

where there is now no problem as to the significance of the quantities which appear, since all refer to molecules of the only gas present.

As examples, the following are the values of the diffusion coefficient in a number of gases at atmospheric pressure and 0°C (in cm²/sec):

Hydrogen–oxygen mixture	0.70
Carbon dioxide–air mixture	0.14
Water vapour–air mixture	0.23
Nitrogen (self-diffusion)	0.18
Oxygen (self-diffusion)	0.18
Carbon dioxide (self-diffusion)	0.10

Diffusion in gases occurs much more rapidly than in liquids. For comparison we may mention that the diffusion coefficient of sugar in water (at room temperature) is only 0.3×10^{-5} cm²/sec, and that of sodium chloride in water is 1.1×10^{-5} cm²/sec.

It is of interest to compare the true distance travelled by gas molecules in their thermal motion with their mean directed displacement in diffusion. For instance, air molecules under

normal conditions travel distances of the order of 5×10^4 cm/sec. The diffusion displacement per second is, in order of magnitude, only $\sqrt{(Dt)} \sim \sqrt{(0.2 \times 1)} \sim 0.5$ cm.

The determination of the thermal conductivity of a gas does not require any essentially new calculations: we need only make use of the analogy noted in §109 between the processes of thermal conduction and diffusion, whereby thermal conduction appears as a “diffusion of energy”, with the thermal diffusivity χ acting as the diffusion coefficient. In a gas, the two processes occur by the same mechanism, namely direct transport by gas molecules. We can therefore say that, in order of magnitude, the thermal diffusivity χ is equal to the self-diffusion coefficient of the gas, i.e.

$$\chi \sim vl.$$

The thermal conductivity κ is obtained by multiplying χ by the specific heat of the gas per unit volume. This volume contains n/N_0 gram-molecules of the gas (where N_0 is Avogadro’s number), and the volume specific heat is therefore nC/N_0 , where C is the molar specific heat; there is no need to distinguish between C_p and C_v , since they are the same in order of magnitude. Thus

$$\kappa \sim \chi nC/N_0 \sim vl nC/N_0,$$

and substituting $l \sim 1/n\sigma$ we have finally

$$\kappa \sim vC/\sigma N_0.$$

The molar specific heat of a gas is independent of its density. We therefore arrive at a remarkable (and at first sight paradoxical) result: the thermal conductivity of a gas depends only on its temperature, and not on its density or pressure.

The specific heat of a gas depends only slightly on the temperature, and the same is true of the cross-section. We can therefore suppose that the thermal conductivity of a gas, like the thermal velocity v , is proportional to \sqrt{T} . In reality, the thermal conductivity increases somewhat more rapidly with temperature, because the specific heat usually increases and the cross-section usually decreases.

As examples, the following are the values of the thermal conductivity of some gases at 0°C (in J/cm.sec.deg):

Chlorine	0.72×10^{-4}	Air	2.41×10^{-4}
Carbon dioxide	1.45×10^{-4}	Hydrogen	16.8×10^{-4}

§114. Mobility

Let consider a gas containing a number of charged particles (ions). If this gas is placed in an electric field, an ordered motion in the direction of the field is superposed on the random thermal motion of the ions which they execute in common with the gas molecules. If the ions were completely free particles, they would move with steadily increasing velocity under the action of the applied field. In reality, however, the ions move freely only in the intervals between collisions with the other particles in the gas. In collisions the particles are randomly scattered, and so the ions essentially lose the directed velocity which they acquire between collisions. Thus a motion results in which the ions, on average, slowly move or *drift* in the direction of the field at a certain velocity u proportional to the field strength.

The order of magnitude of this velocity is easily estimated as follows. An ion of charge e and mass m in an electric field E is subject to a force $F = eE$, which gives the ion an acceleration $w = F/m$. The ion moves with this acceleration during the mean free time τ , and acquires a directed velocity of the order $u \sim w\tau$. Putting $\tau \sim l/v$, where v is the velocity of thermal motion of the ions, we have

$$u \sim Fl/mv.$$

The *drift velocity* u acquired by the ions under the action of the external field is usually written in the form

$$u = KF;$$

the coefficient of proportionality K between the velocity and the force F acting on the ions is called the *ion mobility*.

As examples, the values of the mobility at 20°C and atmospheric pressure are

for H_2^+ ions in H_2 gas	8.6×10^{12} cm/sec.dyn,
for N_2^+ ions in N_2 gas	1.7×10^{12} cm/sec.dyn.

This means, for example, that a field of 1 V/cm will cause N_2^+ ions in nitrogen to drift with a velocity $1.7 \times 10^{12} \times 4.8 \times 10^{-10} \times 1/300 = 3$ cm/sec.

From the above estimate of the velocity u it is seen that $K \sim l/mv$. Comparing this with the diffusion coefficient for the same particles (ions) in the gas, $D \sim lv$, we see that $D \sim mv^2K$, and since $mv^2 \sim kT$ we have

$$D \sim kTK.$$

We shall show that this relation between the diffusion coefficient and the mobility of the particles is in fact an exact equality.

According to Boltzmann's formula, in a state of thermal equilibrium the ion concentration in a gas in a constant external electric field (which we take to be in the direction of the x axis) is proportional to

$$e^{-U(x)/kT},$$

where $U(x) = -Fx$ is the potential energy of an ion in the field; it varies through the gas, increasing in the direction of the field. When a concentration gradient is present, however, a diffusion flux $j = -D dc/dx$ occurs. Let the concentration c be defined as the number of ions per unit volume of the gas, in the form

$$c = \text{constant} \times e^{Fx/kT},$$

since $dc/dx = (F/kT)c$, we have

$$j = -cDF/kT.$$

In a steady (equilibrium) state, however, there can be no transfer of material in the gas. Thus the diffusion flux j in the opposite direction to the field must just compensate the drift flux of the ions in the direction of the field, which is evidently $cu = cKF$. Equating the two expressions, we obtain

$$D = kTK.$$

This relation between the mobility and the diffusion coefficient, called *Einstein's relation*, has been derived here for gases, but is

in fact general. It applies to any particles dissolved or suspended in a gas or a liquid and moving under the action of any external field (electric or gravitational).

§115. Thermal diffusion

In discussing diffusion in a gas mixture, we have so far tacitly assumed that the temperature (and pressure) of the gas is everywhere the same, so that diffusion occurs only because of the concentration gradient in the mixture. In reality it is found that a temperature gradient also may bring about diffusion. In a non-uniformly heated mixture, diffusion occurs even if the composition is uniform; the difference in the thermal motion of the molecules of different components of the mixture (i.e. the difference in their thermal velocities and cross-sections) has the result that the two components appear in different proportions in the numbers of molecules crossing any area in the direction of the temperature gradient and in the opposite direction. The occurrence of a diffusion flux under the action of a temperature gradient is called *thermal diffusion*. This phenomenon is particularly important in gases, which we shall henceforward consider, but exists in principle in liquid mixtures also.

The diffusion flux j_T in thermal diffusion is proportional to the temperature gradient in the gas, and is customarily written in the form

$$j_T = D_T \frac{1}{T} \frac{dT}{dx}$$

The quantity D_T is called the *thermal diffusion coefficient*. Here we should specify exactly what is meant by the flux j_T (unlike the case of ordinary diffusion, where the coefficient D is independent of the way in which the flux is defined); we shall not pause to do this, however. Whereas the diffusion coefficient D is always positive, the sign of the thermal diffusion coefficient is by its nature indeterminate, depending on which component of the mixture is considered.

When the concentration of either component of a mixture tends to zero, the thermal diffusion coefficient must become zero, since there is of course no thermal diffusion in a pure gas. Thus the thermal diffusion coefficient depends considerably on the

concentration of the mixture, again unlike the ordinary diffusion coefficient.

Because of thermal diffusion, concentration differences occur between regions at different temperatures even in a gas mixture of initially uniform composition. These concentration differences in turn cause ordinary diffusion, which acts in the opposite direction, i.e. tends to annul the concentration gradient that has been formed. Under steady conditions, when a constant temperature gradient is maintained in the gas, these two opposite effects finally bring about a steady state in which the two fluxes compensate each other; in this state there is a certain difference in composition between the "hot" and "cold" ends of the gas.

Let us consider the simple case where the molecules of the two gases in the mixture are so different in mass that the thermal velocity of the "heavy" molecules is small compared with that of the "light" molecules. The light molecules, on colliding with the heavy molecules, which may be regarded as at rest, rebound from them elastically, and under these conditions we need consider only the diffusion transport of the lighter component of the mixture.

Let n_1 be the number of molecules of the light component per unit volume, and v_1 their thermal velocity. The flux of this component in the x direction can be estimated by taking the difference between the values of the product $n_1 v_1$ at the points $x - l_1$ and $x + l_1$, where l_1 is the mean free path of the molecules. As in §113, this difference may be replaced by

$$-l_1 d(n_1 v_1)/dx.$$

Hence we see that the transfer of material ceases (i.e. a steady state is established) when the product $n_1 v_1$ becomes constant throughout the gas. But $n_1 = cn$, where c is the concentration of the light component, and n the total number of molecules per unit volume, which is equal to p/kT . Since the total pressure p of the gas is everywhere the same, and the thermal velocity v_1 is proportional to \sqrt{T} , the condition that $n_1 v_1$ is constant implies that the ratio c/\sqrt{T} is constant. In other words, in the steady state the concentration of the light component is greater in the hotter regions.

This is in fact the way in which the composition varies in the majority of cases: the lighter gas usually accumulates at the "hot"

end. This rule is not completely general, however, and the mass of the molecules is not the only factor which determines the direction of thermal diffusion.

The phenomenon of thermal diffusion is utilised for the separation of gas mixtures, and in particular for the separation of isotopes. The principle of the method is clear from the construction of a simple "separating column" operating by thermal diffusion (Fig. 130). This consists of a long vertical glass tube with an electrically heated wire along its axis; the walls of the tube are cooled. The hot gas mixture rises along the axis and the cold

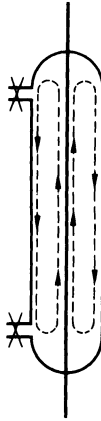


FIG. 130.

mixture descends along the walls. At the same time a process of radial thermal diffusion occurs, as a result of which one component of the mixture (usually that of greater molecular weight) diffuses predominantly to the periphery, and the other to the axis. The components are entrained by the descending and rising currents, and accumulate at the bottom and top of the tube respectively.

§116. Diffusion in solids

Diffusion can also occur in solids, but is an extremely slow process. The phenomenon can be observed, for example, by fusing gold on the end of a rod of lead and keeping it at a high temperature, say 300°C ; even in 24 hours, the gold penetrates about a centimetre into the lead.

There is, of course, also self-diffusion in solids—the mutual diffusion of isotopes of the same substance. This can be observed by means of radioactive isotopes. If, for example, a quantity of a radioisotope of copper is placed on the end of a copper rod and the rod is later cut into pieces, the radioactivity of the pieces gives an idea of the diffusion of the isotope.

The slowness of diffusion in solids is entirely understandable in view of the nature of the thermal motion of the atoms in them. In gases, and even in liquids, the random thermal motion of the molecules includes a “translational component”, the molecules moving through the volume occupied by the body. In solids, however, the atoms are almost always near certain equilibrium positions (the lattice points) and execute small oscillations about these; such a motion can not lead to any general movement of the atoms through the body, nor therefore to diffusion. Only atoms which leave their positions in the lattice and move to other lattice points can take part in diffusion.

However, each atom in a solid is surrounded by a potential barrier. An atom can leave its position only by surmounting this barrier, and to do so it must have sufficient energy. A similar situation has been discussed in connection with the rates of chemical reactions (§91), where we saw that the number of molecules able to react is proportional to an “activation factor” of the form

$$e^{-E/RT}.$$

The number of atoms which can participate in diffusion will also be proportional to such a factor, and therefore so will the diffusion coefficient. The value of the activation energy E per atom (E/N_0) is usually between a fraction of an electron-volt and several electron-volts. For example, in the diffusion of carbon in iron E is about 100 kJ/mole (i.e. about 1 eV per atom); for the self-diffusion of copper, E is about 200 kJ/mole (about 2 eV per atom).

Thus the diffusion coefficient in solids increases very rapidly with increasing temperature. For instance, the diffusion coefficient of zinc in copper increases by a factor of 10^{14} when the temperature is raised from room temperature to 300°C . One of the most rapidly diffusing pairs of metals is gold and lead, which have already been mentioned above. The diffusion coefficient of

gold in lead at room temperature is 4×10^{-10} cm²/sec; at 300°C it is 1×10^{-5} cm²/sec. These figures also show the slowness of the diffusion process in solids.

The acceleration of diffusion by increasing the temperature is the basis of the *annealing* of metals: in order to make the composition of an alloy homogeneous, it is held for a considerable time at a high temperature. The same method is used to relax internal stresses in metals.

In solid solutions of the interstitial type, the solute atoms occupy positions in the "gaps" between atoms at the original lattice points. The diffusion in such solutions (e.g. of carbon in iron) takes place simply by the movement of the interstitial atoms from one gap to another. In substitution-type solutions, however, in an ideal crystal, all the available places are occupied; diffusion in such an ideal crystal would have to take place by simultaneous exchange of positions of two different atoms. In an actual crystal, there are always unoccupied places or vacancies, as already mentioned in §105. These play an important part in the actual mechanism of diffusion, which occurs by atoms from adjoining occupied lattice points "jumping" to the vacant positions.

CHAPTER XV
VISCOSITY

§117. **The coefficient of viscosity**

Let us consider a flow of liquid (or gas) in which the velocity of flow is different at different points. This is not an equilibrium state, and processes will occur which tend to equalise the velocities of flow. Such processes are called *internal friction* or *viscosity*. Just as there is a heat flux from the hotter to the colder parts of a medium in thermal conduction, so in internal friction the thermal motion of the molecules causes a transfer of momentum from the faster to the slower regions of the flow.

Thus the three phenomena of diffusion, thermal conduction and viscosity have analogous mechanisms. In all three there is an equalisation of a property of the body (composition, temperature, or velocity of flow) if this property is originally not uniform through the body; this brings about an approach to a state of thermal equilibrium. In all three cases this is achieved by a molecular transport of some quantity from one part of the body to another. In diffusion there is a transport of number of particles of the various components of the mixture, in thermal conduction a transport of energy, and in internal friction a transport of momentum. All these effects are therefore often combined under the general name of *transport phenomena*.

Let us suppose that a liquid is flowing in the same direction at all points, i.e. that the flow velocity vector \mathbf{u} has the same direction throughout the flow, and suppose also that the magnitude u of the velocity varies in only one direction, perpendicular to that of the velocity, the direction of its variation being taken as that of the x axis: $u = u(x)$.

By analogy with the diffusion flux and the heat flux, we can define the momentum flux Π as being the total momentum transported per unit time in the positive direction of the x axis across unit area perpendicular to that axis. In exactly the same way as

for the other transport processes, we can say that the momentum flux is proportional to the gradient of the flow velocity u :

$$\Pi = -\eta \, du/dx.$$

The quantity η is called the *coefficient of viscosity* or simply the *viscosity* of the medium.

The dimensions of the flux Π are those of momentum divided by area and time, i.e. $[\Pi] = \text{g/cm}\cdot\text{sec}^2$. The dimensions of du/dx are 1/sec. Hence

$$[\eta] = \text{g/sec}\cdot\text{cm}.$$

The unit of viscosity in the CGS system is the *poise* (P).

The viscosity determines the rate of transport of momentum from one point in the flow to another. The velocity is equal to the momentum divided by the mass. The rate of equalisation of the flow velocity is therefore determined by the quantity η/ρ , where ρ is the density, i.e. the mass of the liquid per unit volume. The quantity $\nu = \eta/\rho$ is called the *kinematic viscosity*, whereas η itself is called the *dynamic viscosity*. It is easily seen that

$$[\nu] = \text{cm}^2/\text{sec},$$

i.e. ν has the same dimensions as the diffusion coefficient and the thermal diffusivity; the kinematic viscosity is a kind of diffusion coefficient for velocity.

Let us suppose that a liquid flows in contact with a solid surface; for example, along the walls of a pipe. Between the surface of a solid and any actual liquid (or gas) there always exist forces of molecular cohesion which have the result that the layer of liquid immediately adjoining the surface is entirely brought to rest and “adheres” to the surface. Thus the flow velocity is zero at the wall, and increases away from the wall into the liquid; as a result of viscosity, there then occurs a flux of momentum from the liquid towards the wall.

As we know from mechanics, the change in the momentum of a body per unit time is the force acting on the body. The momentum Π transported through unit area per unit time and ultimately transferred from the liquid to the wall represents the frictional

force exerted on unit area of the solid wall by the liquid flowing past it.

The following comment should be added concerning the simple formula for Π given above. Although the formal analogy already mentioned exists between the phenomena of diffusion, thermal conduction and viscosity, there is also an important difference between them, due to the fact that concentration and temperature are scalar quantities, whereas velocity is a vector. Here we have taken only the simple case where the velocity is everywhere in the same direction; the formula given above for Π is valid only in this case. The impossibility of applying this formula when the direction of the velocity \mathbf{u} is different at different points is evident from the example of a liquid rotating uniformly as a rigid body together with a cylindrical vessel, about the axis of the vessel. The circular velocity of the liquid particles increases with distance from the axis. There is, nevertheless, no flux of momentum, i.e. no frictional forces, in the liquid; a uniform rigid rotation of the liquid (if there is no friction in the suspension of the vessel) does not affect thermal equilibrium and could continue indefinitely without the velocity's becoming uniform.

§118. Viscosity of gases and liquids

The viscosity of a gas may be estimated from the fact that internal friction, thermal conduction and self-diffusion all occur in a gas by the same molecular mechanism. In this case the quantity analogous to the diffusion coefficient is the kinematic viscosity $\nu = \eta/\rho$. We can therefore say that, for a gas, all three quantities ν , χ and D are of the same order of magnitude, and thus we have $\nu \sim \nu l$. The gas density $p = nm$, where m is the mass of a molecule and n the number of molecules per unit volume; hence we have for the viscosity $\eta = \nu\rho$ the expression

$$\eta \sim mn\nu l \sim mv/\sigma,$$

where σ is the collision cross-section.

We see that the viscosity, like the thermal conductivity, is independent of the pressure of the gas. Since the thermal velocity v is proportional to \sqrt{T} , we may suppose that the viscosity of the gas is also proportional to the square root of the temperature. This conclusion is, however, valid only if we can regard the

collision cross-section σ as constant. It has been mentioned in §112 that the cross-section increases somewhat with decreasing temperature. Accordingly, the viscosity decreases with decreasing temperature more rapidly than \sqrt{T} .

The extent to which the approximate equality of the coefficients ν , χ and D is maintained for gases can be seen, for example, from their values for air at 0°C: the kinematic viscosity $\nu = 0.13$, the thermal diffusivity $\chi = 0.19$, and the self-diffusion coefficient of nitrogen and oxygen $D = 0.18$.

The following are the values of the viscosities of some gases and liquids at 20°C:

	η (g/sec.cm)	ν (cm ² /sec)
Hydrogen	0.88×10^{-4}	0.95
Air	1.8×10^{-4}	0.150
Benzene	0.65	0.72
Water	0.010	0.010
Mercury	0.0155	0.0014
Glycerine	15.0	12.0

It is interesting to note that, whereas the dynamic viscosity of water is considerably greater than that of air, the reverse is true for the kinematic viscosity.

The viscosity of a liquid usually decreases with increasing temperature; this is reasonable, since the relative motion of the molecules becomes easier. In liquids of low viscosity, such as water, the decrease is appreciable but not very great. There are, however, liquids, mainly organic (such as glycerine), whose viscosity decreases very rapidly with rising temperature. For example, a rise in temperature of 10° (from 20 to 30°C) causes the viscosity η of water to decrease only by 20%, whereas that of glycerine decreases by a factor of 2.5. The decrease in viscosity of such liquids takes place exponentially, in proportion to a factor of the form $e^{-E/RT}$; for glycerine, $E \approx 65\,000$ J/mole. As we already know (cf. §116), this law of temperature dependence signifies that the occurrence of the process (in this case the relative motion of the molecules) requires the overcoming of a potential barrier.

When the temperature decreases, a viscous liquid rapidly congeals into an amorphous solid. It has already been men-

tioned in §52 that the difference between a liquid and an amorphous solid is purely quantitative. For example, rosin is a solid at room temperature, but even at 50–70°C it behaves as a fluid of high but measurable viscosity, 10^6 to 10^4 P; for comparison we may note that the consistency of honey or syrup corresponds to a viscosity of about 5×10^3 P.

The mechanical properties of liquids such as glycerine and rosin are interesting in another respect also. The characteristic difference between a solid and a liquid is that the solid resists a change in shape (has a shear modulus) but the liquid does not. We may say that the molecular structure of a liquid is instantly “adjusted” to a change in shape; in typical liquids, this occurs in a time of the order of the periods of thermal vibration of the molecules (10^{-10} to 10^{-12} sec). In liquid rosin, however, this “adjustment” requires a longer time, and when the deformation varies very rapidly it may be unable to occur; in rosin at 50–70°C the characteristic time is 10^{-4} to 10^{-5} sec. Thus such a substance behaves as an elastic solid with a certain shear modulus under a very rapidly changing external action (due to sound waves for example), but with respect to a slowly varying action it behaves as an ordinary liquid with a certain viscosity.

§119. Poiseuille's formula

We may use the formula $\Pi = -\eta \, du/dx$ to solve a number of simple problems relating to the flow of a viscous liquid.

Let us first calculate the frictional force between two parallel solid planes in relative motion, with a liquid of viscosity η in the space between them. Let u_0 be the velocity of this motion,

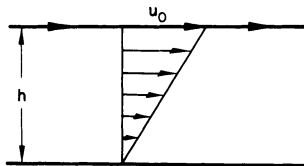


FIG. 131.

and h the distance between the planes; in Fig. 131 the lower plane is at rest and the upper plane moves with velocity u_0 . The liquid adjoining the walls is carried along by them, so that the velocity of the liquid is zero and u_0 at the lower and upper walls

respectively. In the region between the walls, the velocity u varies linearly:

$$u = (u_0/h)x,$$

where x is the distance from the lower wall; this result is derived in the same way as in the exactly similar problem of thermal conduction in a plane layer (§110). The required frictional force acting on unit area of each of the solid planes and tending to slow down their relative motion is given by the momentum flux Π , as described in §117; this is

$$\Pi = u_0\eta/h,$$

and is thus proportional to the relative velocity u_0 of the planes and inversely proportional to the distance between them.

Let us next consider the flow of liquid in a cylindrical tube of radius a and length L , with different pressures p_1 and p_2 maintained at the ends of the tube; the liquid then flows along the tube under the action of the pressure difference $\Delta p = p_2 - p_1$. The flow velocity u of the liquid is everywhere along the axis of the tube, and its magnitude varies in the radial direction (perpendicular to the axis), depending on only one coordinate, the distance r from the axis. We can therefore write the momentum flux transported radially as

$$\Pi = -\eta \, du/dr.$$

Let us consider a volume of liquid bounded by a cylindrical surface of radius r within the tube and coaxial with it. The total flux of momentum through this surface (whose area is $2\pi rL$) is

$$2\pi rL\Pi = -2\pi rL\eta \, du/dr.$$

This is the frictional force exerted on the volume of liquid in question by the remaining liquid, and is balanced by the force due to the pressure difference between the ends of the cylinder, which is $\pi r^2\Delta p$. Equating these forces, we obtain

$$du/dr = -(r/2L\eta) \Delta p,$$

whence

$$u = -(r^2/4L\eta) \Delta p + \text{constant}.$$

The arbitrary constant is determined from the condition that the velocity is zero at the surface of the tube, i.e. for $r = a$. Thus we have finally

$$u = (\Delta p/4L\eta)(a^2 - r^2).$$

Thus a liquid flowing in a tube has what is called a parabolic velocity profile: the velocity varies quadratically from zero at the wall to a maximum value ($u_{\max} = a^2\Delta p/4L\eta$) on the axis of the tube (Fig. 132).

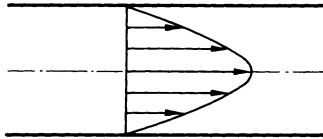


FIG. 132.

Let us determine the mass M of liquid leaving the tube per unit time. If $V(r)$ denotes the volume of liquid leaving per unit time through the cylinder of radius r , the differential of this function is evidently

$$dV(r) = u(r) dS,$$

where $u(r)$ is the velocity of the liquid at a distance r from the axis, and dS is the area of an annulus of radius r and width dr . Since $dS = 2\pi r dr$, we have

$$\begin{aligned} dV(r) &= 2\pi r u dr \\ &= (\pi\Delta p/2L\eta)(a^2 - r^2)r dr \\ &= (\pi\Delta p/4L\eta)(a^2 - r^2)d(r^2). \end{aligned}$$

Hence

$$V(r) = (\pi\Delta p/4L\eta)(a^2r^2 - \frac{1}{2}r^4);$$

the arbitrary constant is taken as zero, since we must have $V(0) = 0$. The total volume of liquid leaving the tube per unit

time is equal to the value of $V(r)$ for $r = a$. Multiplying this by the density ρ of the liquid, we find the required mass:

$$M = (\pi\Delta p/8L\nu)a^4.$$

This is called *Poiseuille's formula*. We see that the quantity of liquid leaving the tube is proportional to the fourth power of the tube radius.

The examples discussed above relate to steady flow of a liquid, in which the velocity of the liquid at any point in the flow is constant in time. One example of non-steady motion may be mentioned here. Let us assume that a disc immersed in a liquid executes torsional oscillations in its plane; the liquid entrained by the disc also oscillates. These oscillations are, however, damped with increasing distance from the disc, and the question arises of the order of magnitude of the distance at which an appreciable damping occurs. This question is formally equivalent to the one discussed in §111 concerning thermal oscillations caused by a plate with variable temperature. The required "penetration depth" L of the oscillatory motion in the liquid is obtained by replacing the thermal diffusivity χ , in the formula derived in §111, by the kinematic viscosity ν of the liquid:

$$L \sim \sqrt{(\nu/\omega)},$$

where ω is the frequency of the oscillations.

§120. The similarity method

Some simple problems of the motion of a liquid have been discussed above. In more complex cases, an exact solution of the problem usually involves very great mathematical difficulties, and is as a rule impossible. For example, it is not possible to give a general solution of the motion through a liquid of a body having even such an apparently simple form as that of a sphere.

Consequently, in various problems of the motion of a liquid, great importance attaches to simple methods based on consideration of the dimensions of the physical quantities on which the motion can depend.

Let us consider, for example, the uniform motion of a solid sphere through a liquid, and let the problem be to determine the

drag force F experienced by the sphere. [Instead of regarding the body as moving through a liquid, we could take the precisely equivalent problem of flow of a liquid past a body at rest; this statement of the problem corresponds to observations of gas flow past bodies in a wind tunnel.]

The physical properties of a liquid which determine its flow or the motion of bodies in it are described by only two quantities: the density ρ and the viscosity η . In addition, in the case considered the motion depends on the velocity u of the sphere and its radius a .

Thus we have altogether four parameters, whose dimensions are as follows:

$$[\rho] = \text{g/cm}^3, \quad [\eta] = \text{g/cm}\cdot\text{sec}, \quad [u] = \text{cm/sec}, \quad [a] = \text{cm}.$$

From these we can form a dimensionless quantity as follows. The dimension g, first of all, can be eliminated in only one way: by dividing η by ρ to give the ratio $\nu = \eta/\rho$, with dimensions $[\nu] = \text{cm}^2/\text{sec}$. Next, to eliminate the dimension sec, we divide u by ν : $[u/\nu] = 1/\text{cm}$. A dimensionless quantity is then obtained by multiplying the ratio u/ν by the radius a . This quantity is denoted by the symbol Re :

$$\text{Re} = ua/\nu = \rho ua/\eta,$$

and is called the *Reynolds number*; it is a very important property of the motion of a liquid. Clearly any other dimensionless quantity can only be a function of the Reynolds number.

Let us return now to the determination of the drag force. Its dimensions are $\text{g}\cdot\text{cm}/\text{sec}^2$. A quantity having these dimensions which can be formed from the same parameters is, for example, $\rho u^2 a^2$. Any other quantity having the same dimensions can be written as a product of $\rho u^2 a^2$ and some function of the dimensionless Reynolds number. We can therefore say that the required drag force is given by a formula of the type

$$F = \rho u^2 a^2 f(\text{Re}).$$

The unknown function $f(\text{Re})$ cannot, of course, be determined from dimensional considerations alone, but we see that by means of these considerations we have been able to reduce the

problem of determining a function of four parameters (the force F as a function of ρ , η , u and a) to a problem of determining a single function $f(\text{Re})$. This function may be found experimentally, for example. By measuring the drag on any one sphere in any one liquid and plotting from the results a graph of the function $f(\text{Re})$, we can find the drag in the motion of any sphere in any liquid.

The above arguments are general and apply, of course, to steady motion in a liquid of bodies of any form (not only spherical). The quantity a in the Reynolds number must then be taken as some linear dimension for a body of given shape, and we are thus able to compare the flow round geometrically similar bodies which differ only in size.

Motions which have the same value of the Reynolds number for different values of the parameters ρ , η , u , a are said to be *similar*. The entire pattern of the motion of the liquid in such cases differs only in the scale of distances, velocities etc.

Although for brevity we have spoken of liquids, the whole of the above discussion applies to gases. The only condition assumed to be satisfied is that the density of the medium (liquid or gas) does not undergo any appreciable change during the motion, and may therefore be regarded as constant; in such circumstances the moving medium is said to be *incompressible*. Although from the ordinary point of view a gas is an easily compressible medium, the changes in pressure which occur in a gas during its motion are usually insufficient to cause any appreciable change in its density. The gas ceases to behave as an incompressible medium only at velocities comparable with that of sound.

§121. Stokes' formula

Let us again consider the drag F encountered by a body moving in a liquid (or gas). When the velocity is sufficiently small, the drag force is always proportional to the velocity. In order to derive such a relation from the formula

$$F = \rho u^2 a^2 f(\text{Re}),$$

we must suppose that at low velocities the function $f(\text{Re})$ is of the form constant/Re. This gives

$$F = \text{constant} \times \eta au.$$

We see that, if the drag is proportional to the velocity, it necessarily follows that the drag is also proportional to the linear dimension of the body (and to the viscosity of the liquid).

The determination of the proportionality coefficient in this relation requires more detailed calculations. When a sphere moves in a liquid, the constant is found to be 6π , i.e.

$$F = 6\pi\eta au,$$

where a is the radius of the sphere. This is *Stokes' formula*.

The above discussion enables us to state more precisely what is meant by a "sufficiently small" velocity for Stokes' formula to be valid. Since the form of the function $f(\text{Re})$ is in question, the required condition must relate to the values of the Reynolds number, and since the number Re is proportional to the velocity u (for a given size of the body), it is clear that the condition for the velocity to be small must be that the dimensionless number Re is small:

$$\text{Re} = au/v \ll 1.$$

Hence we see that the condition for the velocity to be "sufficiently small" is a relative one. The actual range of permissible velocities depends on the size of the moving body (and on the viscosity of the liquid). For very small bodies (e.g. fine particles, suspended in a liquid, in Brownian motion) Stokes' formula is valid even for velocities which in other respects could not be regarded as small.

If a sphere moves in a liquid under the action of an external force P (for instance, the force of gravity, with allowance for the partial loss of weight in the liquid), a uniform motion will finally be established with a velocity such that the force P is just balanced by the drag force. If $P = F$, the velocity is given by

$$u = P/6\pi a\eta.$$

This formula is frequently used to determine the viscosity of a liquid from a measurement of the rate of fall of a solid sphere in it. The viscosity may also be determined by means of Poiseuille's

formula, by measuring the rate of outflow of a liquid from a pipe along which it is impelled by a given pressure difference.

Stokes' formula is also the basis of a method of measuring the unit charge, first used by Millikan to measure the charge on the electron. In these experiments fine droplets produced by an oil spray were placed in the space between horizontal plates forming a plane capacitor. The droplets have a charge owing to electrification in the spraying process or absorption of ions from the air. By observing under a microscope the rate of fall of a droplet by the effect of its weight alone, we can use Stokes' formula to calculate the radius and hence the mass of the drop (whose density is known). Then, by applying a suitable potential difference across the capacitor, we can bring the droplet to rest, the downward force of gravity being balanced by the upward electrical force on the charged droplet. Knowing the weight of the droplet and the electric field strength, we can calculate the charge on the droplet. Such measurements show that the charge on the droplets is always an integral multiple of a certain quantity, which is evidently the unit charge.

§122. Turbulence

The flow of a liquid in a pipe as described in § 119 is orderly and smooth: each liquid particle moves in a fixed straight line and the whole pattern of the flow is like the motion of various layers of liquid with different relative velocities. This kind of regular steady flow of a liquid is called *laminar* flow.

It is found, however, that a liquid flow of such a kind continues to occur only when the Reynolds number is sufficiently small. For flow in a pipe, this number may be defined by the formula $Re = ud/\nu$, where d is the diameter of the pipe and u the mean velocity of the liquid. For example, if the flow velocity is increased in a pipe of given diameter, a point is reached at which the nature of the flow changes completely. It becomes extremely disordered, and instead of smooth lines the liquid particles describe tangled, meandering and continually changing paths. Such motion is said to be *turbulent*.

The difference between the two types of motion appears very clearly if we observe the flow in a glass tube and introduce a small quantity of a coloured liquid into the flow through a capillary. At low velocities, the coloured liquid is carried along by the

main flow as a thin straight filament. At high velocities, however, this filament is disrupted and the coloured liquid mixes rapidly and almost uniformly with the entire flow.

If we follow the variation of the liquid velocity with time at any given point in a turbulent flow, we find irregular random fluctuations of the velocity about some mean value. The mean values of the velocity describe the pattern of motion of the liquid in which the irregular turbulent fluctuations or *eddies* are smoothed out. This averaged velocity is usually what is meant in speaking simply of the velocity of a turbulent liquid flow.

Turbulent mixing of a liquid is a much more efficient means of transferring momentum than the process of molecular transfer by internal friction in a laminar flow. For this reason the velocity profile over the cross-section of a pipe in turbulent flow is considerably different from that in laminar flow. In the latter, the velocity gradually increases from the wall to the axis of the pipe, but in turbulent flow the velocity is almost constant over a large part of the cross-section of the pipe, falling rapidly in a thin layer adjoining the wall to the value zero which it must have at the wall itself.

The unimportant role of viscosity in comparison with turbulent mixing also has more general consequences: the viscosity has no direct effect on the properties of turbulent flow, and these properties are therefore determined by a smaller number of quantities than in laminar flow, since these do not include the viscosity of the liquid. The possibilities of constructing different combinations having the same dimensions from the remaining quantities are much more restricted, and the application of the similarity method may therefore immediately give more specific results.

Let us find, for example, the relation between the mean velocity u of flow in a pipe and the pressure gradient which brings about this flow (i.e. the ratio $\Delta p/L$, where Δp is the pressure difference between the ends of the pipe and L the length of the pipe). The quantity $\Delta p/L$ has the dimensions $\text{g.cm}^{-2}.\text{sec}^{-2}$. The only combination having these dimensions which can be constructed from the available quantities (the velocity u , the diameter d of the pipe and the density ρ of the liquid) is $\rho u^2/d$. We can therefore assert that

$$\Delta p/L = \text{constant} \times \rho u^2/d,$$

where the constant is a pure number. Thus in turbulent flow in a pipe the pressure gradient is proportional to the square of the mean velocity, and not to the velocity itself as in laminar flow. [This law, however, is only approximately valid, since no account has been taken of the effect of the boundary layer at the wall, in which the velocity decreases very rapidly and the viscosity plays an important part.]

It has already been mentioned that flow in a pipe becomes turbulent for sufficiently large Reynolds numbers. Experiment shows that, for this to occur, the Reynolds number must exceed 1700. For smaller values of the Reynolds number, laminar flow is completely stable. This means that, when the flow is perturbed by some external agency (vibration of the pipe, roughness of the pipe inlet, etc.), the resulting deviations from smooth flow are rapidly damped. When $Re > 1700$, on the other hand, perturbations of the flow lead to disruption of the laminar condition and the appearance of turbulence. By means of special precautions to reduce the perturbations which inevitably occur, it is possible to postpone the transition to turbulent flow until still higher values of Re are reached, and laminar flow in a pipe has been observed even for $Re = 50\,000$.

Turbulence is a general feature of flow at very high Reynolds numbers. It occurs not only in flow in a pipe but also in flow of a liquid (or gas) past various solid bodies (or, what amounts to the same thing, motion of these bodies through a liquid). Let us consider the pattern of such flow in more detail.

Because of the law of similarity discussed in §120, it is immaterial what is the precise reason for the large value of the Reynolds number, whether the size a of the body is large, the velocity u is large, or the viscosity η is small. In this sense we can say that for very large Reynolds numbers the liquid behaves as if its viscosity were very low. This applies, however, only to a liquid flowing far from solid walls. Near the surface of a solid a thin *boundary layer* is formed in which the velocity decreases from the value corresponding to frictionless motion to the value zero corresponding to the adhesion of the viscous liquid to the wall. The thickness of the boundary layer decreases with increasing Reynolds number. Within this layer the velocity changes very rapidly, and hence the viscosity of the liquid is of decisive importance.

The properties of the boundary layer lead to the important phenomenon of *separation* in flow past bodies. When a liquid flows past a body, it first of all moves along the front part of the body, which becomes wider in the direction of flow. The streams of liquid undergo a kind of compression, and their velocity accordingly increases and the pressure decreases, as follows from Bernoulli's equation (see §61). In flow along the rear part of the body, however, the streams expand as the body narrows, the velocity in them decreases, and the pressure correspondingly rises. Thus in this part of the flow the pressure increases in the direction of flow, i.e. the pressure difference opposing the motion of the liquid increases. This pressure difference arising in the main flow acts also on the liquid in the boundary layer and retards it. The liquid particles in the boundary layer move more slowly than in the main flow and begin to move even more slowly, until, when a point on the surface of the body is reached at which the pressure is sufficiently high, the particles in the boundary layer come to rest and begin to move in the opposite direction. Thus a reversed motion occurs near the surface of the body, despite the fact that the main flow continues to move in the same direction. At points still further along the surface of the body, the reverse flow becomes wider and ultimately displaces the main flow completely, which thus becomes separated from the surface.

This motion with a reversed flow, however, is entirely unstable and immediately becomes turbulent. The turbulence extends forward along the flow and gives rise to a long strip of liquid in turbulent motion behind the body, called the *turbulent wake*; this is shown diagrammatically in Fig. 133. For example, this

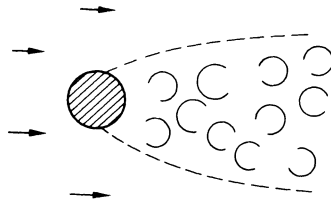


FIG. 133.

occurs for a sphere when Re is greater than about 1000, Re being defined as du/ν , where d is the diameter of the sphere.

At very large Reynolds numbers the formation of the turbulent wake is the principal cause of the drag on the body moving in the liquid. Under these conditions we can again use dimensional arguments to determine the law of drag. The drag force F on a body (of given shape) can depend only on the size a of the body, its velocity u and the density ρ of the liquid, but not on its viscosity. From these three quantities only one combination having the dimensions of force can be derived, namely the product $\rho u^2 a^2$. We can therefore say that

$$F = \text{constant} \times \rho u^2 a^2,$$

where the constant depends on the shape of the body. Thus, at very high Reynolds numbers the drag force is proportional to the square of the velocity (*Newton's law of drag*). It is also proportional to the square of the linear size of the body or, what is the same thing, to its cross-sectional area (which is itself proportional to a^2). Finally, the drag force is proportional to the density of the liquid. In the opposite case of small Reynolds numbers, it will be remembered that the drag is proportional to the viscosity of the liquid and independent of the density. Whereas at small Reynolds numbers the drag is determined by the viscosity of the liquid, when Re is large the effect of the inertia (mass) of the liquid becomes predominant.

The drag at large Reynolds numbers depends very greatly on the shape of the body, which determines the point of separation of the flow and therefore the width of the turbulent wake. The drag due to the turbulent wake is smaller when the wake is narrower. This determines the choice of shape of a body such that the drag force is the least possible; such a shape is said to be *streamlined*.

A streamlined body must be rounded at the front and drawn out at the back to run smoothly to a pointed end, as shown in Fig. 134 (this diagram may be regarded as showing the longitudinal cross-section of an elongated solid of revolution or as the cross-section of a "wing" of large span). The liquid flow along

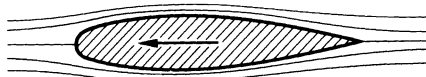


FIG. 134.

such a body closes up smoothly at the pointed end, with no sharp turns anywhere; this eliminates the rapid rise of pressure in the direction of the flow. The flow is separated only at the actual point, and the turbulent wake is therefore extremely narrow.

In connection with the drag at high velocities, it should be mentioned that the foregoing discussion refers only to velocities small compared with that of sound, so that the liquid may be regarded as incompressible.

§123. Rarefied gases

The conclusions drawn in §§113 and 118 concerning transport processes in gases are valid only so long as the gas is not too rarefied: the mean free path of the molecules must be small compared with the size of the bodies considered (the vessel containing the gas, the bodies moving through the gas, etc.). However, even at a pressure of 10^{-3} to 10^{-4} mm Hg the mean free path is 10–100 cm, and is comparable with or even exceeds the usual dimensions of apparatus. A similar situation occurs in problems of space flight near the earth: even at an altitude of about 100 km the mean free paths of the particles in the ionised gas are several tens of metres.

Here we shall apply the term *rarefied* to gases in which the mean free path of the molecules is large compared with the size of the bodies. This criterion depends not only on the state of the gas itself but also on the size of the bodies that are in question. A given gas may therefore behave as a rarefied gas in some conditions but not in others.

Let us consider the transfer of heat between two solid plates heated to different temperatures and immersed in a gas. The mechanism of this process is totally different in rarefied and non-rarefied gases. In non-rarefied gases, the transfer of heat from the hotter to the colder wall occurs by the gradual “diffusion of energy” transferred from one molecule to another in collisions. But if the mean free path l of the gas molecules is large in comparison with the distance h between the walls, the molecules in the space between the plates will scarcely ever collide with one another, and will move freely after reflection from one plate until they collide with the other. When scattered from the hotter plate, the molecules gain some energy from it, and they give up part of their energy on colliding with the colder plate.

Under these conditions there is obviously no meaning in speaking of the temperature gradient in the gas between the plates, but by analogy with the expression $q = -\kappa dT/dx$ for the heat flux we can now define the "thermal conductivity" of a rarefied gas by the relation

$$q = -\kappa(T_2 - T_1)/h,$$

where $T_2 - T_1$ is the temperature difference between the plates. We can estimate this conductivity in order of magnitude directly by analogy with the expression derived in §113 for the ordinary thermal conductivity:

$$\kappa \sim vlnC/N_0.$$

There is no need to repeat the derivation; we need only note that, since we now have collisions with the plates instead of between molecules, the mean free path in this formula must be replaced by the distance h between the plates:

$$\kappa \sim vhnC/N_0$$

(C is the molar specific heat of the gas, v the thermal velocity of the molecules, and n the number of molecules per unit volume). Substituting $n = p/kT$ and replacing the product N_0k by the gas constant R , we have

$$\kappa \sim phvC/RT.$$

We see that the "thermal conductivity" of a rarefied gas is proportional to its pressure, unlike the thermal conductivity of a non-rarefied gas, which is independent of the pressure. It must be emphasised, however, that this conductivity is no longer a quantity pertaining to the gas alone: it depends also on the distance h between the two bodies.

The decrease in the thermal conductivity of a rarefied gas with decreasing pressure is the basis of the use of an evacuated space for thermal insulation, as for instance in Dewar vessels for keeping liquefied gases; these have double walls with the air evacuated from the space between. As the evacuation proceeds the ther-

mal conductivity of the air at first remains constant, and begins to decrease rapidly only when the mean free path becomes comparable with the distance between the walls.

Internal friction in rarefied gases behaves similarly. Let us consider, for example, two solid surfaces with a layer of rarefied gas between them and moving with relative velocity u . The “viscosity” of the gas is defined by the relation

$$\Pi = \eta u/h,$$

where Π is the frictional force per unit area acting on the solid surfaces, and h the distance between them. Replacing the mean free path l by h in the formula $\eta \sim nmvl$ derived in §118, we obtain

$$\eta \sim nmvh,$$

and putting $n = p/kT$ and $kT \sim mv^2$ we have finally

$$\eta \sim ph/v.$$

Thus the “viscosity” of a rarefied gas is also proportional to the pressure. Like the thermal conductivity, the viscosity depends not only on the properties of the gas itself but also on the characteristic dimensions which appear in a given problem.

Let us use the above expression for η to estimate the drag F on a body moving in a rarefied gas. Here h must be taken as the linear size a of the body. The frictional force per unit area of the surface of the body is

$$\Pi \sim \eta u/a \sim pu/v,$$

where u is the velocity of the body. Multiplying this by the area S of the surface of the body, we obtain

$$F \sim upS/v.$$

Thus the drag exerted by a rarefied gas is proportional to the surface area of the body, unlike the drag in a non-rarefied gas, which is proportional to the linear size of the body.

We may also mention some interesting effects relating to the outflow of a rarefied gas through small holes, of size much less than the mean free path of the molecules. This outflow, called *effusion*, is quite different from the ordinary outflow through large holes, where the gas flows out in a jet like a continuous medium. In effusion the molecules leave the vessel independently and form a "molecular beam" in which each molecule moves with the velocity with which it approached the hole.

The rate of outflow of a gas in effusion, i.e. the number of molecules leaving the hole per unit time, is in order of magnitude Snv , where S is the area of the aperture. Since $n = p/kT$, $v \sim \sqrt{(kT/m)}$, we have

$$Snv \sim Sp/\sqrt{(mkT)}.$$

The rate of effusion thus decreases with increasing mass of the molecule. Thus, in effusion of a mixture of two gases, the outflowing gas will be enriched in the lighter component. This is the basis of one common method of isotope separation.

Let us now imagine two vessels containing gases at different temperatures T_1 and T_2 and connected by a small hole (or a tube of small diameter). If the gases were not rarefied, the pressures in the two vessels would become equal and the force exerted by each gas on the other at the hole would be the same. For rarefied gases, however, this is no longer true, since the molecules pass freely through the hole without colliding with one another. The pressures p_1 and p_2 then take values such that the numbers of molecules passing through the hole in each direction are the same. According to the expression derived above for the rate of outflow, this means that the condition

$$p_1/\sqrt{T_1} = p_2/\sqrt{T_2}$$

must hold. Thus different pressures are set up in the two vessels, the pressure being higher in the vessel with the higher temperature. This is called the *Knudsen effect*. In particular, it must be taken into account in the measurement of very low pressures; a difference in the temperatures of the gas under examination and the gas in the measuring instrument causes a corresponding difference in pressure.

§124. Superfluidity

It has already been mentioned that liquid helium is exceptional in its physical properties, being a “quantum liquid”, whose properties cannot be understood on the basis of the concepts of classical mechanics. This is shown by the fact that helium remains liquid at all temperatures down to absolute zero (§72).

Helium becomes liquid at 4.2°K . At about 2.2°K it remains liquid but undergoes a further transformation, a phase transition of the second kind (see §74). At temperatures above the transformation point liquid helium is generally called helium I, and below it helium II.

Helium II has the following properties. Firstly, heat transfer is extremely rapid. A temperature difference between the ends of a capillary filled with helium II is very quickly eliminated, and helium II is in fact the best heat conductor known. This property, incidentally, explains the striking change that is seen on observing visually the transformation of helium I into helium II: the surface of the continuously boiling liquid suddenly becomes completely calm and smooth when the transition point is reached. The reason is that, because of the very rapid removal of heat from the vessel walls, the vapour bubbles typical of boiling are no longer formed there, and helium II evaporates only from its free surface.

The fundamental and primary property of liquid helium, however, is that of *superfluidity*, discovered by P. L. Kapitza. This refers to the viscosity of liquid helium.

The viscosity of a liquid can be measured from its rate of flow through narrow capillaries, but in the present case this method is unsuitable, and a method is required which allows the flow of a greater quantity of liquid than can pass through a narrow capillary. This is achieved by an experiment in which helium II passes along a very narrow gap (about 0.5μ wide) between two discs of ground glass (Fig. 135). Yet even under these conditions liquid helium exhibits no viscosity whatever, which shows that the viscosity is exactly zero. The absence of viscosity in helium II is what is referred to by the term “superfluidity”.

The superfluidity of helium II is directly responsible for the “creeping film” which it forms. When helium II is placed in two vessels separated by a partition, it spontaneously takes the same level in both vessels after a certain time. This transfer

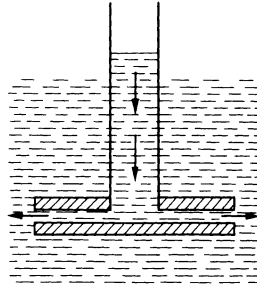


FIG. 135.

takes place along a thin film (a few hundred ångströms thick) which is formed by the helium on the walls and which acts as a siphon (Fig. 136). The mere fact that a film is formed is not a specific property of helium II. A film is formed by any liquid which wets the solid surface. In ordinary liquids, however, the formation of the film and its spreading over the surface occur extremely slowly because of the viscosity of the liquid. The

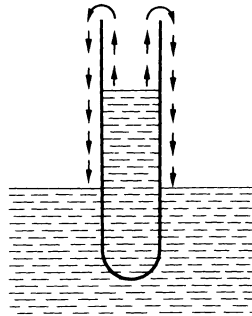


FIG. 136.

formation and movement of the film occur rapidly for helium II, however, because of its superfluidity. The velocity of the creeping film is several tens of centimetres per second.

We have discussed the viscosity of helium as measured by the rate of flow of the liquid through a narrow gap. But the viscosity of a liquid can also be measured in another way. If a disc (or cylinder) suspended in the liquid executes torsional oscillations about its axis, the friction on the disc which retards its oscillations is a measure of the viscosity. It is found that in

such measurements helium II shows a small but non-zero viscosity (of the order of 10^{-5} poise).

The theory (due to L. D. Landau) which gives an explanation of these paradoxical properties of liquid helium cannot be explained here before the principles of quantum mechanics have been stated. We shall, however, describe the remarkable physical picture which results from this theory.

It usually appears self-evident that, in order to describe the motion of a liquid, it is sufficient to specify its velocity at every point. Yet this seemingly obvious statement is invalid for the motion of the quantum liquid helium II.

Helium II is found to be capable of executing two motions simultaneously, so that, in order to describe the flow, the values of not one but two velocities at every point must be specified. The situation may be visualised by regarding helium II as a mixture of two liquid components which can move independently “through” each other without mutual friction. In reality, however, there is only one liquid, and it must be emphasised that the “two-fluid” model of helium II is only a convenient way of describing the phenomena which occur in it. Like any description of quantum phenomena in classical terms, it is not completely adequate—as is reasonable, since our intuitive ideas reflect what we encounter in ordinary life, whereas quantum phenomena usually appear only in the microscopic world inaccessible to our direct perception.

Each of the two motions simultaneously occurring in liquid helium involves the displacement of a certain mass of liquid. In this sense we can speak of the densities of the two “components” of helium II, although it must again be emphasised that this terminology does not at all signify that the atoms of the substance can actually be divided into two classes. Each of the two motions is a collective motion of a large number of the same liquid atoms.

The two motions have entirely different properties. One motion takes place as if the corresponding “component” of the liquid had no viscosity; this is called the *superfluid* component. The other (*normal*) component moves like an ordinary viscous liquid.

Another important difference between the two types of motion in helium II is that the normal component transports heat in

its motion, but the superfluid motion is not accompanied by any transport of heat. In a certain sense we may say that the normal component is the heat itself, which in liquid helium becomes independent and separate from the mass of the liquid, and becomes able to move relative to a "background" that is at absolute zero. This picture is radically different from the usual classical idea of heat as a random motion of atoms, inseparable from the whole mass of the substance.

These concepts give an immediate explanation of the principal results of the experiments described above. First of all, they eliminate the contradiction between the measurements of the viscosity of the liquid from the friction experienced by a rotating disc and from the flow of liquid through a gap. In the first case the disc is retarded because, when it rotates in liquid helium, it undergoes friction with the "normal" part of the liquid, and the viscosity of this component is essentially what is measured. In the second case, the superfluid part of the helium flows through the gap, while the normal component, which has a viscosity, is retarded and "percolates" very slowly through the gap. Thus, in this experiment, the zero viscosity of the superfluid component is observed.

But since the superfluid motion does not transfer heat, the outflow of helium through the gap causes a kind of filtering off of liquid without heat, the heat remaining in the vessel. In the ideal limiting case of a sufficiently narrow gap, the liquid leaving would be at absolute zero. In an actual experiment its temperature is not zero, but is lower than that of the vessel. For example, by driving helium II through a porous filter its temperature can be lowered by 0.3 to 0.4°, and this is a large amount when the temperature is only 1–2°K.

The very high rate of heat transfer in helium II also finds a natural explanation. Instead of the slow process of molecular transport of energy in ordinary thermal conduction, we here have a rapid process of heat transfer by the normal component of the liquid. The relation between the process of heat transfer in helium II and the occurrence of motion in it is clearly shown by the following experiment. A light vane is placed in front of a hole in a small vessel filled with liquid helium (and immersed in liquid helium) (Fig. 137). When the helium in the vessel is heated, the vane is deflected. This occurs because heat flows

from the vessel as a stream of the viscous normal component, which deflects the vane in front of the hole. In the opposite direction there is an inflow of the superfluid component, so that the actual quantity of liquid in the vessel is unchanged and the latter remains full. The superfluid component has no viscosity and does not move the vane as it flows past it.

The presence of two “components” in helium II is shown directly in an experiment based on the following idea. When a cylindrical vessel containing liquid helium rotates, only part of the liquid should be carried round, namely the “normal” part, which undergoes friction against the walls; the “superfluid”

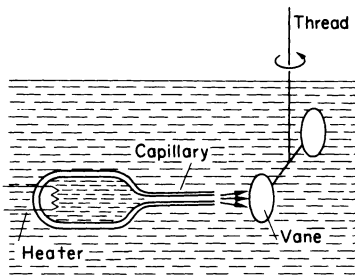


FIG. 137.

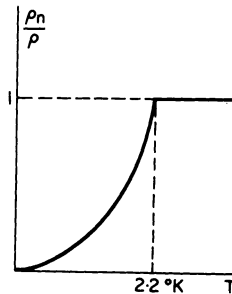


FIG. 138.

part should remain at rest. In an actual experiment, the rotation of the vessel is replaced by torsional oscillations of a stack of numerous thin discs; this increases the area of the surface which entrains the liquid.

At temperatures above the transition point the liquid (helium I) is entirely in the normal state and is entirely carried round by the rotating walls. At the transition point a qualitatively new property of the liquid first appears, namely the occurrence of the superfluid component, and this is the nature of the phase transition of the second kind in liquid helium. As the temperature decreases further, the fraction of the superfluid component becomes greater, and at absolute zero the liquid should become entirely superfluid. Figure 138 shows the form of the temperature dependence of the ratio of the density ρ_n of the normal component of the liquid helium to the total density ρ of the liquid (the sum of the normal and superfluid densities ρ_n and ρ_s is, of course, always equal to the total density ρ).

Finally, let us consider a phenomenon in liquid helium related to the propagation of sound waves in it. Sound waves in an ordinary liquid consist of periodic compressions and rarefactions propagated through the liquid, in which each particle executes an oscillatory motion about the mean equilibrium position with a periodically varying velocity. In helium II, however, two different motions can take place simultaneously with different velocities. There are therefore two quite distinct possibilities for the motion in a sound wave. If the two components of the liquid execute an oscillatory motion in the same direction, moving as it were in unison, we have a sound wave of the same kind as in an ordinary liquid.

The two components may also, however, execute oscillations in opposite directions, moving “through” each other, so that the masses transported in each direction are almost exactly equal. In such a wave (called a *second sound* wave) there are almost no compressions and rarefactions of the liquid as a whole. Instead, there are periodic oscillations of temperature in the liquid, since the mutual oscillations of the normal and superfluid components are essentially oscillations of heat relative to the “superfluid background”. Thus the second sound wave is a kind of “thermal wave”, and it is therefore clear that the source needed to create such a wave is a heater whose temperature varies periodically.

The whole of the foregoing discussion has referred simply to liquid helium. It is necessary to make explicit that this discussion relates only to one of the isotopes of helium, namely the common isotope He^4 . There exists also another much rarer isotope, He^3 . By the methods of nuclear physics it is possible to separate this isotope in quantities sufficient for liquefaction and experiments with the liquid. It is again a “quantum liquid”, but its properties are entirely different, and in particular it is not a superfluid. Although the two isotopes of helium are chemically identical, there is a very important difference between them, due to the fact that the nucleus of the He^4 atom comprises an even number of particles (protons and neutrons), and that of He^3 an odd number. This difference has the result that the quantum properties of the two substances are entirely different, and so in turn brings about the difference in the physical properties of the liquids which they form.

INDEX

- Absolute temperature 145
Absolute zero 147
Acceleration 10
 angular 74
 in circle 11–12
 dimensions of 17
 due to gravity 60
 normal 12
 tangential 12
 units of 17
Action and reaction, law of 14
Activation energy 268
Activation factor 268
Adiabatic processes 178–82
Adsorption 279–82
Aggregate states of matter 152, 197
Allotropy 218
Amplitude of oscillation 87
Angular acceleration 74
Angular frequency 87
Angular momentum 36
 conservation of 37, 40
 rotational 72–73
Angular velocity of rigid body 66–68
Anisotropy 132
Atomic number 106
Atomic weight 105, 110
Atoms 105–11
Avogadro's law 155
Avogadro's number 105
Axis of symmetry 115
 rotary-reflection 116
 screw 130
Azeotropic mixture 243
- Barometric formula 159
Beats 102
Bending of rods 297–8
Bernoulli's equation 186
Boiling 206–7, 241–6
- Boltzmann factor 163
Boltzmann's constant 146
Boltzmann's formula 158
Boyle's law 156
Bravais lattices 121–9
Brownian motion 144
Buffer solution 266
Bulk modulus 300
- Capillarity effects 276, 285–9
Carnot cycle 191
Catalysis 270
Celsius scale of temperature 147
Central field, motion in 40–43
Centre of gravity 78
Centre of inversion 115
Centre of mass 9, 78
 frame of reference 10
Centre of symmetry 115
Centrifugal energy 83
Centrifugal force 83
CGS system of units 18
CGSE system of units 45, 50
Chain reactions 272–5
Charge, *see* Electric charge
Charles' law 157
Chemical elements 105
 lattices of 134–7
Chemical equilibrium 254–6
Chemical reactions 252–75
Classical mechanics 5
Clausius–Clapeyron equation 203
Close packing 135
Closed system 5
Collisions 31
 elastic 31–36
 inelastic 31
Colloidal solutions 291–3
Compressibility 174, 300
Compression of solids 294, 298–301

- Conduction, thermal 319–34
 Conductivity, thermal 320, 333–4, 358
 Conservation
 of angular momentum 37, 40
 of charge 47
 of energy 24
 of mass 8
 of momentum 6
 Conserved quantities 6
 Contact, angle of 282–5
 Convection 319
 Coriolis force 83
 Corresponding states, law of 215
 Cosmic velocity
 first 62
 second 64
 Coulomb's law 44
 Couple 78
 arm of 78
 moment of 78
 Critical point 207–9
 of mixing 234
 Critical pressure 207, 209
 Critical temperature 207, 209
 of mixing 234
 Cross-section 329
 Crystal
 classes 132
 lattices 120–43
 modifications 218–21
 planes 139–42
 space groups 129–31
 systems 123–9
 Crystals
 biaxial 129
 defects in 308–12
 liquid 227–9
 mixed 234–6
 ordering of 225–7
 single 153
 structure of 120–43
 uniaxial 129
 Cyclic process 167

 Dalton's law 157
 Damping coefficient 95
 Defects in crystals 308–12
 Degrees of freedom 3, 68
 Derived units 15

 Diffusion 318
 coefficient 319
 thermal 336
 flux 318
 in gases 330–3
 in solids 338–40
 thermal 336–8
 Diffusivity, thermal 320
 Dimensions 17–18, 25
 Dislocations 309–12
 Dissociation
 constant 264
 degree of 264
 electrolytic 262
 Drag force 348–51, 359
 Dulong and Petit's law 176

 Efficiency of heat engine 191–2
 Effusion 360
 Elastic
 deformations 306
 energy 297
 limit 306
 properties 296
 stresses 294
 Electric
 charge 44
 conservation of 47
 dimensions of 45
 unit 106
 units of 45
 field 46–56
 potential of 49
 flux 51
 lines of force 48
 tubes of force 52
 Electrical interaction 45–56
 Electrolytes
 strong 262–3
 weak 264–6
 Electrons 106–8
 Electrostatic
 potential 49–50
 units 45
 Enantiomorphism 131
 Endothermic reaction 252
 Energy
 centrifugal 83
 conservation of 24
 dimensions of 25

- Energy (*cont.*)
 elastic 297
 internal 26–27
 of particle 24
 of rigid body 68–72
 total 24
 units of 25–26
 see also Kinetic energy; Potential energy
 Enthalpy 170
 Entropy 194–6
 Equalisation time 326–7
 Equation of motion
 of particle 13
 of rotating rigid body 74
 Equation of state 150
 for ideal gas 155
 Equilibrium constant 257
 Equipotential surface 50
 Equivalence, principle of 61
 Eutectic point 247
 Exothermic reaction 252
 Extension of solids 294–7
- Field, *see* Force field
 Finite motion 29
 Fluctuations 194
 Flux, electric 51
 Force 13
 centrifugal 83
 Coriolis 83
 dimensions of 17
 field 19, 44–65
 electric 46–56
 gravitational 56–65
 inertia 81–85
 restoring 88
 resultant 77
 units of 17, 18
 Foucault pendulum 84
 Fractional distillation 244
 Frame of reference 1
 centre-of-mass 10
 inertial 2
 laboratory 34
 Free motion 1
 Freedom, degrees of 3, 68
 Frequency of motion 86, 87
 angular 87
 Friction 94, 316–17
 internal 341
- Gas
 constant 156
 ideal 153–60
 rarefied 357–60
 Gauss' theorem 51, 59
 Glide plane 130
 Gravitation, Newton's law of 57
 Gravitational
 constant 57
 field 59–65
 interaction 56–65
 potential 58
 Gravity 60–62
 centre of 78
 Gyroscope 78–81
- Hardening 306, 314
 Harmonic oscillations 86–93
 Heat
 content 170
 function 170
 of reaction 252–4
 of solution 230
 of transition 201
 see also Thermal; Thermodynamics
 Helium, liquid 200, 218, 225, 361–6
 Henry's law 232
 Hooke's law 295
- Ideal gas 153–60
 Incompressible medium 350
 Indices of crystal planes 141–3
 Inertia
 forces 81–85
 law of 2
 moment of 69–71
 Inertial frame 2
 Infinite motion 29
 Internal energy 26–27
 Internal motion 26
 Inversion
 centre of 115
 point 184
 Ionisation potential 108
 Irreversibility of thermal processes
 188–9, 192–4
 Isomorphism 235
 Isotopes 109–11
- Joule-Kelvin process 182–4

- Kelvin scale of temperature 147
- Keplerian motion 63
- Kepler's
 - first law 64
 - second law 41
 - third law 63
- Kinetic energy
 - of particle 24
 - of rigid body 68–70
 - and temperature 145
- Kinetic processes 318
- Knudsen effect 360

- Laboratory frame of reference 34
- Laminar flow 352
- Latent heat 201
- Le Chatelier's principle 170
- Length, units of 15–16
- Lever rule 200
- Lines of force 48
- Liquid crystals 227–9
- Loschmidt's number 155

- Macroscopic properties 131, 150
- Mass
 - centre of 9, 78
 - conservation of 8
 - number 109
 - of particle 6
 - reduced 41
 - of system 10
 - units of 7, 16–17, 58
- Mass action, law of 257
- Maxwellian distribution 163
- Mean free path 328
- Mechanical properties of solids 294–317
- Meniscus 283
- Mixed crystals 234–6
- Mixing of liquids 232–4, 246–9
- Mobility 334
- Molecularity of reactions 270–2
- Molecules 111–14
 - symmetry of 117–20
- Moment
 - arm 37
 - of couple 78
 - of inertia 69–71
- Moments, law of 75
- Momentum
 - angular, *see* Angular momentum
 - of closed system 6
 - conservation of 6
 - of particle 6
- Monomolecular layer 280

- Nernst's theorem 177
- Newtonian mechanics 5
- Newton's
 - first law 2
 - law of drag 356
 - law of gravitation 57
 - second law 13
 - third law 15
- Nucleus of atom 106

- One-dimensional motion 27
- Optical isomers 119
- Ordering of crystals 225–7
- Oscillations 86–104
 - damped 93–96
 - forced 96–102
 - natural 89
 - small 88–93
 - torsional 93
- Osmosis 236–8
- Osmotic pressure 236–8

- Parametric resonance 102
- Particle 3
- Pascal's law 294
- Pendulum
 - compound 91
 - Foucault 84
 - simple 90
 - simple equivalent 92
- Periodic motion 29, 86–104
- Perpetual motion 189
- Phase
 - diagrams 198–200
 - of matter 197
 - of oscillation 87
 - rule 251
 - transitions 197–229
 - of second kind 222–5
- Physical system of units 18
- Plane of symmetry 115
 - glide 130
- Plastic deformations 306, 312–15
- Poiseuille's formula 348

- Poisson's
 adiabatic 181
 ratio 296
 Polymorphism 218
 Potential barrier 30
 Potential energy 22
 of electric charges 45
 of electric field 49
 of gravitational field 58
 in uniform field 23
 Potential well 30
 Power 26
 Precession 79
 Pressure 149–51
 osmotic 236–8
 partial 157
- Quantity of heat 168
- Radius vector 3
 Raoult's law 240
 Rarefied gases 357–60
 Reactive forces 7–8
 Relativity of motion 2
 Resonance 98–100
 parametric 102
 Restoring force 88
 Resultant force 77
 Reverse condensation 246
 Reversibility of motion 187
 Reynolds number 349–56
 Rigid body 66
 motion of 66–85
 Rigidity, modulus of 302
 Rotary-reflection axis 116
 Rotation of rigid body 66–68, 72–76
- Saturated
 solution 230
 vapour 203–7
- Screw axis 130
- Sectorial velocity 41
- Self-diffusion 332
- Semipermeable membranes 236
- Separated flow 355
- Shear deformations 301–5
- SI system of units 18, 45, 50
- Similar motions 350
- Solid solutions 234–6
- Solidification of liquid mixtures 241–9
- Solubility 231
 product 263
- Solute 231
- Solutions 231–51
 dilute 231
 solid 234–6
 weak 231, 237
- Solvent 231
- Space groups 129–31
- Specific heat 169–70
 of gases 171–4
 of solids and liquids 176–7
- Statistical weight 194
- Steady flow of gases and liquids 184–7
- Stereoisomerism 119
- Stokes' formula 351
- Sublimation 217
- Supercooling of vapours 207, 289–91
- Superfluidity 361–6
- Superheating of liquids 207, 290–1
- Superposition of fields 46
- Supersaturation of vapours 207, 289–91
- Surface
 -active substances 280
 energy 276
 phenomena 276–93
 pressure 285
 tension 276–9, 281–2
- Symmetry 115–43
 axis of 115, 116, 130
 centre of 115
 elements 115–16
 plane of 115, 130
 translational 120
- Temperature 144–9
- Thermal
 conduction 319–34
 conductivity 320, 333–4, 358
 diffusion 336–8
 diffusivity 320
 equilibrium 144, 151, 188–9, 193–4, 254
 expansion 175–6
 motion 144
 processes 178–96
 resistance 324
 velocity 147
- Thermodynamic temperature 145

- Thermodynamics
 - first law of 168
 - second law of 189
- Time, units of 16
- Top 80–81
- Torque 38
- Torques, law of 75
- Torricelli's formula 187
- Torsion of rods 303–5
- Translation
 - of crystal lattice 120
 - of rigid body 66
- Transport phenomena 341
- Triple point 216
- Tube of force 52
- Turbulent flow 352–7
- Two-body problem 41

- Uniform compression 300
- Uniform deformation 297, 298–301
- Uniform extension 300
- Uniform field
 - motion in 19–20
 - potential energy in 23
- Unilateral compression 301
- Unit cell 120
- Units 15–18
 - CGS system 18
 - CGSE system 45, 50
 - derived 15
 - physical system 18
 - SI system 18, 45, 50

- Van der Waals
 - equation 211
 - interaction 113–14
- Velocity 4
 - addition rule 5
 - angular, *see* Angular velocity
 - cosmic 62, 64
 - dimensions of 17
 - of escape 64
 - sectorial 41
 - thermal 147
 - units of 17
- Viscosity 341–66
 - dynamic 342
 - kinematic 342

- Weight 59
- Wetting of surfaces 283–5
- Work 20–26, 166–7
 - dimensions of 25–26
 - units of 25–26
- Work-hardening 306, 314

- Yield point 308
- Young's modulus 296

- Zero-point vibrations 148–9, 173

